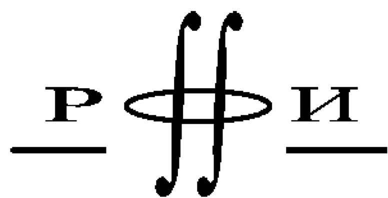
VISIT…

LANZAROTE Caliente.COM

Математические модели нелинейной динамики.

А.И.Чуличков

УДК 519.8
ББК 22.161
Ч 89



Издание осуществлено при поддержке Российского фонда фундаментальных исследований по проекту 00-01-14091

Чуличков А.И. **Математические модели нелинейной динамики.** — М.: ФИЗМАТЛИТ, 2000. — 296 с. — ISBN 5-9221-0083-1.

Обобщаются известные и предлагаются новые методы математического моделирования нелинейных динамических систем. На простых примерах пояснены механизмы возникновения динамического хаоса, самоорганизации и др. Предложен принципиально новый подход к моделированию динамических систем, основанный на теории возможностей и нечеткой математике. Он ориентирован на описание динамики в условиях неопределенности и является альтернативой стохастическому моделированию. Предлагаются методы прогноза динамики на основе наблюдений над системой, выполненных с погрешностью.

Для специалистов по математическому моделированию, а также для студентов старших курсов и аспирантов технических и физико-математических специальностей вузов.

Табл 5. Ил. 91. Библиогр. 20 назв.

ISBN 5-9221-0083-1

© ФИЗМАТЛИТ, 2000
© А.И. Чуличков, 2000

Оглавление

Предисловие 9

Глава 1. **Математические модели динамических систем. Основные определения** 13

1.1. Определение динамической системы 13

1.1.1. Динамическая система и ее состояние (13). 1.1.2. Моделирование динамической системы (14).

1.2. Динамическая система, описываемая конечной системой дифференциальных уравнений 14

1.2.1. Гармонические колебания (15). 1.2.2. Движение в поле потенциальных сил (16). 1.2.3. Нелинейный осциллятор (17). 1.2.4. Консервативные и диссипативные системы (18). 1.2.5. Маятник с затуханием (18). 1.2.6. Нелинейный осциллятор Ван дер Поля (19). 1.2.7. Странные аттракторы (19).

1.3. Дискретные эволюционные модели 20

1.3.1. Разностные эволюционные уравнения (20). 1.3.2. Отображение Пуанкаре (21).

1.4. О решении начальных задач для дифференциальных эволюционных уравнений 21

1.4.1. Область определения фазовых траекторий (21). 1.4.2. Автономные динамические системы (22). 1.4.3. Типы траекторий автономных динамических систем (22). 1.4.4. Предельные точки и предельные множества (22).

Глава 2. **Классификация поведения динамических систем** 24

2.1. Топологическая эквивалентность 24

2.1.1. Определение топологической эквивалентности (24). 2.1.2. Зависимость от параметра (25).

2.2. Исследование качественного поведения систем 27

2.2.1. Примеры влияния управляющих параметров на динамику систем (27). 2.2.2. Грубые динамические системы (29).

2.2.3. Классификация особых точек (32). 2.2.4. Поведение вблизи особых точек (33).
2.3. Устойчивость динамических систем 36
2.3.1. Устойчивость особых точек (36). 2.3.2. Предельные циклы (37). 2.3.3. Устойчивость по Ляпунову (38). 2.3.4. Орбитальная устойчивость (38). 2.3.5. Устойчивость и ляпуновские характеристические показатели (39). 2.3.6. Устойчивость периодических решений (40).
2.4. Типичные бифуркации нелинейных динамических систем 41
2.4.1. Бифуркация смены устойчивости (41). 2.4.2. Бифуркация "седло–узел". Складка (42). 2.4.3. Сборка (42). 2.4.4. Бифуркация рождения предельного цикла (43). 2.4.5. Бифуркации удвоения периода и расщепления цикла (45).
2.5. Введение в элементарную теорию катастроф 46
2.5.1. Вводные замечания и примеры (46). 2.5.2. Неморсовские особые точки. Росток и возмущение катастрофы (48). 2.5.3. Классификация катастроф (49). 2.5.4. Канонический вид эволюционного уравнения в неособой точке (50). 2.5.5. Канонический вид эволюционного уравнения в морсовской точке. Лемма Морса (50). 2.5.6. Канонический вид эволюционного уравнения в особой точке катастроф. Возмущения. (52). 2.5.7. Сепаратрисы на множестве параметров (53). 2.5.8. Флаги катастроф (54). 2.5.9. Фазовый переход как катастрофа (56). 2.5.10. Точки катастроф и изменение климата (58).

Глава 3. **Приближенные методы исследования нелинейных систем** 61
3.1. Метод усреднения 61
3.2. Асимптотические методы малого параметра 63
3.2.1. Регулярные возмущения системы дифференциальных уравнений (64). 3.2.2. Асимптотические решения сингулярно возмущенных дифференциальных уравнений (69).

Глава 4. **Гамильтоновы системы** 76
4.1. Основы вариационного исчисления 76
4.1.1. Основные понятия (76). 4.1.2. Уравнение Эйлера–Лагранжа (77).
4.2. Задачи динамики 78
4.2.1. Движение в центральном поле. Циклические координаты (79). 4.2.2. Законы сохранения и инвариантность гамильтониана (80). 4.2.3. Особенности фазовых портретов гамильтоновых систем (81).
4.3. Вполне интегрируемые системы 83
4.3.1. Скобки Пуассона и первые интегралы (83). 4.3.2. Условно периодическое движение (83). 4.3.3. Резонансные и нерезо-

нансные торы в фазовом пространстве (85). 4.3.4. О теории Колмогорова–Арнольда–Мозера (86).
4.4. Инвариантные торы в негамильтоновых системах 88

Глава 5. **Хаос в динамических системах** 90
5.1. Что такое хаос 90
5.1.1. Хаос: мифология и математика (90). 5.1.2. Хаотические колебания (93). 5.1.3. Аттрактор Лоренца. Вывод уравнений (94). 5.1.4. Анализ системы уравнений Лоренца (97). 5.1.5. Реакция Белоусова–Жаботинского (98). 5.1.6. Хаос и сечение Пуанкаре (98). 5.1.7. Характерные признаки хаоса (99).
5.2. Дискретные отображения 99
5.2.1. Сдвиг Бернулли (99). 5.2.2. Треугольное отображение (101). 5.2.3. Математические характеристики хаоса (104). 5.2.4. Хаотическая диффузия (105).
5.3. Сценарии перехода к хаосу 105
5.3.1. Переход к хаосу через удвоение периода (105). 5.3.2. Переход к хаосу через перемежаемость (110). 5.3.3. Сценарий Рюэля–Такенса (112).
5.4. Эргодичность и перемешивание 113
5.4.1. Среднее по времени и среднее по ансамблю (113). 5.4.2. Эргодические системы (114). 5.4.3. Перемешивающие системы (115). 5.4.4. Диссипативные перемешивающие системы (117). 5.4.5. Странные аттракторы (118). 5.4.6. Фрактальные свойства странного аттрактора (120).
5.5. Ведущие параметры 122
5.5.1. Быстрое и медленное время (122). 5.5.2. Параметры порядка и принцип подчинения (123).

Глава 6. **Фракталы: определения и свойства** 127
6.1. Что такое фрактал 127
6.1.1. Примеры и определения фрактала (127). 6.1.2. Фрактальность пространственных форм (129). 6.1.3. Динамические фракталы (130).
6.2. Размерность Хаусдорфа–Безиковича 131
6.3. Фрактал как самоподобный объект 135
6.3.1. Фракталы Жюлиа и Мандельброта (136). 6.3.2. Фрактальные кластеры (138). 6.3.3. Фракталы как модели физических систем (138).
6.4. Самоподобие как фундаментальное свойство природы 140

Глава 7. **Численные методы исследования динамических систем** 143
7.1. Расчет отображений Пуанкаре 144
7.2. Численный анализ периодических решений 146

7.3. Вычисление спектра ляпуновских характеристических показателей . . . 147
7.4. Расчет размерности аттрактора . . . 151
7.4.1. Оценка топологической размерности (151). 7.4.2. Фрактальная размерность аттрактора (153).

Глава 8. **Самоорганизация в нелинейных системах** . . . 154
8.1. О двух тенденциях динамики — от беспорядка к порядку и обратно . . . 154
8.2. Проблема обратимости времени и ее связь с теорией нелинейных систем . . . 156
8.2.1. Стрела времени и законы динамики (156). 8.2.2. Квантово-механический и космологический парадоксы (159). 8.2.3. Причина необратимости времени в статистической физике (160). 8.2.4. Описание движения, несводимое к траекториям (166).
8.3. Самоорганизация в активных средах . . . 170
8.3.1. Бистабильные среды (171). 8.3.2. Возбудимые среды (176). 8.3.3. Автоколебательные среды (179).
8.4. Нелинейные волны. Солитоны . . . 180
8.4.1. Гиперболические и диспергирующие волны (180). 8.4.2. Солитоны (184).
8.5. Самоорганизация в химической кинетике . . . 187
8.5.1. Реакция Белоусова–Жаботинского (187). 8.5.2. Математическая модель реакций химической кинетики (188). 8.5.3. Брюсселятор (189). 8.5.4. Анализ математической модели брюсселятора (192).
8.6. Самоорганизация в биологических системах . . . 194
8.6.1. Возникновение жизни (194). 8.6.2. Математические модели выживания (195). 8.6.3. Модели роста и взаимодействия популяций (196). 8.6.4. О модели морфогенеза (198).
8.7. Клеточные автоматы . . . 199
8.7.1. Математическая модель клеточного автомата (199). 8.7.2. Игра "Жизнь" (202). 8.7.3. Фильтр движущихся целей (203). 8.7.4. Клеточная модель физической реальности (204).
8.8. Обучающиеся системы . . . 206
8.8.1. Модель Изинга (206). 8.8.2. Нейронные сети (210). 8.8.3. Инвариантные сети (214). 8.8.4. Морфологический анализ изображений (216).

Глава 9. **Системы со случайными шумами** . . . 224
9.1. Роль флуктуаций . . . 224
9.2. Случайные процессы . . . 225

9.2.1. Марковские случайные процессы (226). 9.2.2. Уравнение Смолуховского (227).
9.3. Уравнение для плотности вероятности 228
9.4. Физические системы с шумами 230
9.4.1. Уравнение Ланжевена (230). 9.4.2. Движение в потенциальном поле (231). 9.4.3. Барометрическая формула (232). 9.4.4. Нестационарные решения уравнения Фокера–Планка (233).
9.5. Теория второго порядка 234
9.5.1. Сходимость в среднем квадратичном (234). 9.5.2. Корреляционная функция случайного процесса (234). 9.5.3. Непрерывность в среднем квадратичном (235). 9.5.4. Дифференцируемость и интегрируемость в среднкм квадратичном (235).

Глава 10. **Измерение и прогнозирование** 237
10.1. Теория измерительно-вычислительных систем 237
10.1.1. Несколько неформальных определений (239). 10.1.2. Схема реального измерения (241). 10.1.3. Схема идеального измерения (241). 10.1.4. В чем состоит интерпретация измерения (242). 10.1.5. Линейная модель измерения (242). 10.1.6. Интерпретация измерений с помощью линейных изме- рительно-вычислительных систем (242). 10.1.7. Наблюдения с помощью датчика второго порядка (243).
10.2. Методы синтеза ИВС как идеальных приборов 245
10.2.1. Схема измерений и ее математические модели (245). 10.2.2. Несмещенный синтез выходного сигнала идеального прибора (247). 10.2.3. Синтез идеального прибора с ограничением на энергию шума на его выходе (249). 10.2.4. Синтез выходного сигнала идеального прибора. Априорные данные (252).
10.3. Надежность модели и надежность интерпретации 254
10.3.1. Зачем проверять модель измерения (255). 10.3.2. Надежность модели измерения (256). 10.3.3. Надежность интерпретации (258).

Глава 11. **Нечеткие модели** 260
11.1. Возможность и вероятность 260
11.2. Математические основы теории возможностей 264
11.2.1. Нечеткие множества и события. Алгебра нечетких множеств (264). 11.2.2. Мера возможности и мера необходимости (265).
11.3. Принцип относительности в теории возможностей 268
11.4. Условная мера возможности 270
11.5. Нечеткие элементы 270
11.6. Нечеткое моделирование 272
11.7. Гауссовы нечеткие элементы 273

11.7.1. Определение гауссова нечеткого элемента (273). 11.7.2. Маргинальное распределение (274). 11.7.3. Условное распределение нечеткого гауссова элемента (275). 11.7.4. Аппроксимация нечетких гауссовых элементов (276). 11.7.5. Собственный базис нечеткого гауссова элемента (277).

11.8. Нечеткая динамика 280

11.8.1. Нечеткие процессы (280). 11.8.2. Марковские нечеткие процессы (281). 11.8.3. Уравнение Смолуховского для нечетких процессов (282). 11.8.4. Однородные нечеткие процессы с независимыми приращениями (283). 11.8.5. Процессы с дискретным временем и конечным числом состояний (286). 11.8.6. Волны возможности (286). 11.8.7. Нечеткие процессы и наблюдения (289). 11.8.8. Распространение возможностей (292).

Список литературы 293

Предисловие

Знание фундаментальных законов природы, общества, человека формирует основы мировоззрения, способность видеть проявление основных принципов природы в многообразии закономерностей частных наук, дает подходы, позволяющие действовать в соответствии с этими принципами в новых нестандартных условиях и благодаря этому достигать цели с наименьшими затратами. Целиком взвалить на плечи гуманитарных наук груз формирования мировоззрения было бы не совсем правильно, так как именно концепции естествознания, в основном выдвинутые в XX веке, фактически составляют фундамент современных взглядов на природу и человека в целом, а язык естествознания, главным образом, язык математических моделей, значительно ближе и убедительнее, например, для студентов естественнонаучных специальностей, чем язык философии.

В настоящее время курсы математики, общей и теоретической физики, химии, биологии, геологии и других дисциплин в основном все-таки направлены на формирование конкретных знаний, умения решать конкретные задачи, и тем самым учат студентов говорить каждый на "своем" языке — языке определенного круга явлений и законов. Объединяющую роль мог бы сыграть курс, основанный на широком круге явлений, относящихся не только к физике, курс достаточно математизированный, главное внимание в котором был бы уделено не описанию и подробному изучению феноменов, — это делается на младших курсах, а именно выделению и осознанию общих принципов, лежащих в основе всех этих феноменов, в объединении разнородных явлений мира, формированию общей концепции.

В какой-то мере курс, синтезирующий в одном разные стороны реальности, можно построить на основе современной теории нелинейных динамических систем. Ряд свойств этих систем, такие, как неустойчивость, нелинейность, открытость, диссипация, порождают режимы существования и эволюции, свойственные широкому классу сложных систем, начиная от механических, термодинамических,

химических, и кончая живыми организмами и их сообществами. В первую очередь это хаотические режимы, которые сейчас принято считать характерными этапами развития любой достаточно сложной нелинейной системы, явления самоорганизации, механизм которых может объяснить различные асимметрии физического мира, возникновение жизни, механизмы социальных революций и др. Необычайно высокая восприимчивость систем, находящихся на этапе хаотического развития, дает ключ к пониманию резких скачкообразных переходов, определяет границы предсказуемости их поведения, а также и горизонт реконструкции предшествующих состояний. Анализ сложных нелинейных систем позволяет осознать конструктивную роль кризисов в развитии систем, позволяет найти наилучший стиль поведения (или управления системой) как в период кризисов, так и в период спокойного развития между этапами качественных перестроек.

Попытка создания такого курса привела к появлению этой книги. Она основана на конспекте лекций, читаемых автором студентам пятого курса кафедры компьютерных методов физики физического факультета МГУ.

Первые несколько разделов посвящены математическим методам моделирования нелинейных систем. Это и количественное исследование динамических систем — методы регулярных и сингулярных возмущений дифференциальных уравнений, методы усреднения и др., но в основном — методы качественной теории дифференциальных уравнений: теория устойчивости, классификация особых точек, методы фазового портрета, классификация бифуркаций, основы теории катастроф. Рассматриваются консервативные гамильтоновы системы, приводятся свойства их поведения, в частности, изучаются условно периодические движения, основы теории Колмогорова–Арнольда–Мозера. Большое внимание уделяется наглядным геометрическим образам поведения динамических систем как в зависимости от времени, так и в зависимости от значений внешних управляющих параметров.

Следующая часть монографии посвящена изучению явления динамического хаоса. Первые рассмотренные здесь модели — простейшие одномерные с дискретным временем, такие, как сдвиг Бернулли, треугольное преобразование, логистическое отображение. На их примере демонстрируются универсальные свойства хаоса, в частности, сценарии возникновения хаотических режимов. Далее изучаются и более сложные модели: системы с перемешиванием, как диссипативные, так и консервативные. Формулируется язык описания хаотического поведения, в частности, геометрический язык фракталов.

Заключительная часть монографии состоит из примеров нелинейных систем со сложным поведением. Это нелинейные волны, неравновесные термодинамические системы, модели химической кинетики, популяционной биологии, режимы с обострением. Изучаются

особенности поведения активных сред, модели клеточных автоматов, нейронных сетей. Последние разделы посвящены стохастическому и нечеткому описанию динамики сложных нелинейных систем.

Хотя большинство изучаемых здесь моделей являются классическими для разных естественнонаучных дисциплин, освещение их с единой точки зрения приводит к общей концепции, которую, следуя Г. Хакену, можно назвать синергетической. Автор надеется, что после прочтения книги яснее станут видны общие принципы и в фазовых переходах, и в этапах развития биологических популяций, и в ходе исторического процесса, и в результате модели естествознания оживут, возникнут понятийные связи между дисциплинами, яснее и естественнее проступит единство мира.

Заключительная часть монографии содержит изложение принципиально нового подхода к моделированию динамических систем, основанного на теории возможностей. Этот подход альтернативен к стохастическому и предназначен для создания моделей, описывающих системы в условиях неопределенности. Обычно в таких ситуациях применяются теоретико-вероятностные методы, и при этом многие аспекты неясности и неопределенности моделируются в терминах случайности, отражающей неполноту знаний, их недостоверность, а также нечеткости и неточности, относящихся к их содержанию. Нечеткость и неточность естественно ассоциируются с распределением вероятностей, неясность и неопределенность отражаются в частичном незнании последнего; возникающие в связи с этим проблемы формулируются в терминах теории проверки статистических гипотез и теории оценивания.

Однако теоретико-вероятностные методы оказались неэффективными при моделировании широкого класса процессов и явлений, в организации которых именно неопределенность и нечеткость играют решающую роль. Дело в том, что в теоретико-вероятностной схеме прогнозируется не поведение системы, а частота того или иного ее поведения, связанная с вероятностью, причем предполагается, что частота не изменяется при заданных условиях, что характеризует так называемую стохастическую устойчивость. Но в сложных ситуациях сами условия меняются достаточно быстро и не подлежат оценке, при этом не имеет смысла говорить о частотах событий. В таких ситуациях можно говорить лишь об определенных тенденциях — например, утверждая, что “если тенденция изменений в условиях сохранится, то вероятность событий увеличится”. Значения вероятностей в такой ситуации бессмысленны для прогнозирования состояния системы, и более естественно описывать лишь эти тенденции и предпочтения, причем не в абсолютной, а в ранговой шкале, оценивая возможность того или иного поведения системы. Это позволяет сделать теория возможностей, развитая в работах Ю.П.Пытьева [16], в которой на

множестве всех состояний системы задается мера, показывающая, какие состояния более возможны, а какие — менее и тем самым характеризующая систему в ранговой шкале. Для выбора оптимального поведения системы, например, минимизирующего возможность риска, достаточно описания тенденций в ранговой шкале, поэтому на практике для принятия оптимальных решений ранговое описание системы в терминах возможности адекватно.

В монографии изучаются модели динамических систем на основе введенного понятия нечеткого процесса, исследуются уравнения и модели, являющиеся нечеткими аналогами стохастических процессов.

Издание этой книги стало возможным благодаря финансовой поддержке РФФИ и помощи моих друзей и коллег. Я глубоко благодарен В.Г. Буданову, плодотворные обсуждения с ним проблем нелинейной динамики дали толчок к написанию книги, ее структура и наполнение во многом определились в результате этих дискуссий. Большое влияние на ее содержание оказали и беседы с С.П. Курдюмовым. Разделы, посвященные морфологическому анализу изображений, измерению и прогнозированию, нечетким моделям написаны под влиянием идей моего учителя Ю.П. Пытьева и в тесном взаимодействии с ним. Приношу им мою искреннюю и глубокую благодарность.

А.И. Чуличков

Глава 1

Математические модели динамических систем. Основные определения

1.1. Определение динамической системы

1.1.1. Динамическая система и ее состояние. Под динамической системой будем понимать объект или процесс, для которых однозначно определено понятие состояния как совокупности значений некоторых величин в заданный момент времени и задан оператор, определяющий эволюцию начального состояния во времени.

Например, система материальных точек с заданным потенциалом взаимодействия является типичным примером динамической системы, так как ее состояние полностью определяется значением начальных координат и импульсов всех точек, а эволюцию системы определяют классические уравнения движения (второй закон Ньютона). Более сложным примером является среда (в частности, атмосфера Земли, содержимое химического реактора и др.) в которой происходят процессы тепло- и массопереноса, физические фазовые переходы, химические реакции и т.п. Состояние такой системы в фиксированный момент времени определяется концентрацией фаз, температурой и другими параметрами, задаваемыми в каждой точке среды. Примерами динамических систем могут служить и биологические системы — организм человека или животного, или взаимодействующие популяции. На вычислительные процессы, происходящие в компьютерах или нейронных сетях, социальные процессы в нашей стране и на другие процессы тоже можно смотреть с позиций теории динамических систем.

Если исследователя интересуют состояния динамической системы в любой момент времени на заданном интервале, то говорят о систе-

мах с непрерывным временем и для описания их эволюции используют, как правило, дифференциальные уравнения. Если же для описания поведения системы достаточно знать ее состояние в конечном или счетном числе моментов времени, то говорят о системе с дискретным временем и для описания ее развития используют дискретные отображения, например, разностные уравнения.

Если данные позволяют однозначно указать значения параметров системы в нужные моменты времени, то говорят о детерминированных динамических системах. Однако эволюцию динамических систем не всегда удается определить однозначно — в этих случаях обычно используют описание системы в терминах о вероятности или возможности того или иного состояния. Распространенной моделью динамической системы являются марковские цепи или процессы, графы и др.

1.1.2. Моделирование динамической системы. Будем говорить, что задана математическая модель динамической системы, если заданы параметры (координаты) системы и эволюционный оператор, позволяющий указать состояние динамической системы (значение координат) в последующие моменты времени. При этом речь может идти об однозначном задании состояния системы в заданные моменты времени, либо об определении распределения вероятности или возможности на множестве ее состояний.

Далее не будем делать различий между динамической системой и ее математической моделью.

1.2. Динамическая система, описываемая конечной системой обыкновенных дифференциальных уравнений

Пусть набор чисел $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ в некоторый момент времени описывает состояние динамической системы, и разным наборам $\{x_1, x_2, \ldots, x_N\}$ соответствуют разные состояния. Зададим эволюционный оператор, указав скорость изменения каждого состояния системы:

$$\frac{dx_i}{dt} = \mathbf{F}_i(x_1, x_2, \ldots, x_N), \quad i = 1, \ldots, N; \tag{1.1}$$

здесь $\mathbf{x}$ — точка евклидова пространства $\mathcal{R}_N$, называемого фазовым пространством, $\mathbf{x}$ называется фазовой точкой. Системы вида (1.1), в которых правая часть не зависит явно от времени, называются автономными. Если систему уравнений дополнить начальными условиями $\mathbf{x}(0) = \mathbf{x}_0$, то получим начальную задачу для уравнений (1.1), или

задачу Коши. Ее решение — $\{\mathbf{x}(t),\ t \geq 0\}$, рассматриваемое как множество точек фазового пространства $\mathcal{R}_N$, образует фазовую траекторию. В фазовом пространстве правые части системы уравнений (1.1) задают векторное поле скоростей, сопоставляющее всякой точке $\mathbf{x}$ вектор $\mathbf{F}(\mathbf{x})$, выходящий из точки $\mathbf{x}$; здесь $\mathbf{F}(\mathbf{x}) = \{F_1(x_1, \ldots, x_N), \ldots \ldots, F_N(x_1, \ldots, x_N)\}$. Фазовые траектории и векторное поле скоростей дают наглядное представление о характере поведения системы с течением времени. Множества фазовых траекторий, соответствующие различным начальным условиям, образуют фазовый портрет динамической системы.

1.2.1. Гармонические колебания. Рассмотрим начальную задачу (задачу Коши) для обыкновенного дифференциального уравнения, описывающего движение материальной точки единичной массы, движущейся вдоль прямой под действием возвращающей силы, пропорциональной смещению точки из состояния равновесия $x = 0$:

$$\begin{gathered} \ddot{x} = -x; \\ x(0) = x_0, \quad \dot{x}(0) = v_0; \end{gathered} \tag{1.2}$$

здесь используются обозначения $\dot{x} = dx/dt$, $\ddot{x} = d^2x/dt^2$. Эта начальная задача для дифференциального оператора второго порядка приводится к задаче Коши для системы двух дифференциальных уравнений первого порядка:

$$\begin{gathered} \dot{x} = p, \\ \dot{p} = -x, \\ x(0) = x_0, \quad p(0) = v_0. \end{gathered} \tag{1.3}$$

Значения координаты x материальной точки и ее импульса $p = \dot{x}$ полностью описывает состояние механической системы в момент времени t. Фазовыми координатами в этом случае является пара (x, p). Принято считать, что в такого рода системах число степеней свободы равно единице (поскольку речь идет об одномерном механическом движении), а фазовое пространство двумерно. Решая задачу Коши, имеем описание эволюции системы в виде $x = a\sin(t + \varphi)$, $p = a\cos(t + \varphi)$.

На фазовой плоскости фазовые траектории являются окружностями, радиусы которых определяются начальными условиями (рис. 1.1).

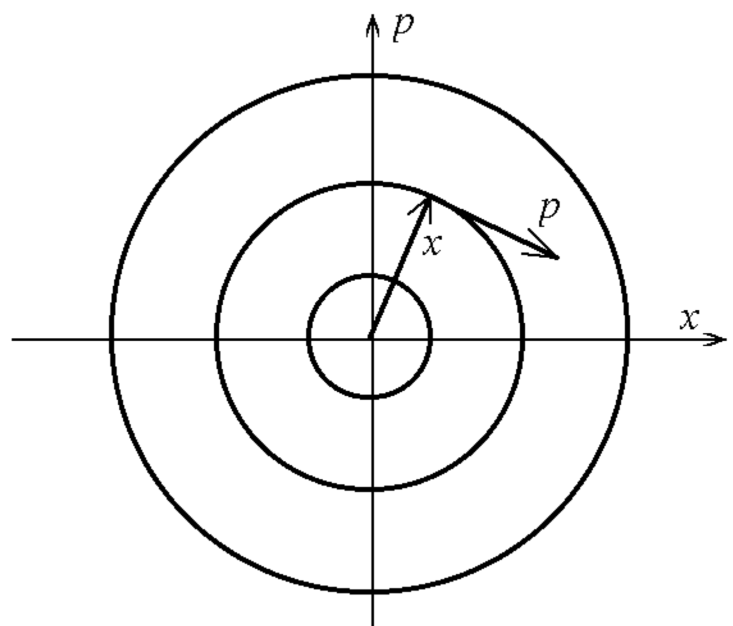


Рис. 1.1. Фазовые траектории математического маятника

Заметим, что, как видно из соотношения (1.3), вектор фазовой скорости $\mathbf{F}(\mathbf{x}) = d\mathbf{x}/dt$ точки $\mathbf{x} = (x, p)$ равен $(p, -x)$, перпендикулярен вектору (x, p) и равен ему по модулю — это обеспечивает равномерное вращение точки $(x(t), p(t))$ в фазовой плоскости с течением времени. Так как кинетическая энергия T колеблющейся материальной точки равна $T = \dot{x}^2/2 = p^2/2$, а потенциальная $U = x^2/2$, то полная энергия системы $E = T + U = \|\mathbf{x}\|^2/2$ пропорциональна квадрату радиуса окружности, и всякому значению энергии соответствует своя фазовая кривая — окружность. Множество окружностей, соответствующих различным значениям полной энергии системы, образуют фазовый портрет.

Заметим, что в этой динамической системе можно ввести новые координаты на фазовой плоскости — полярные координаты (E, φ), в которой семейство фазовых траекторий можно записать в простом виде: $E = \text{const}$.

1.2.2. Движение в поле потенциальных сил. Пусть эволюция системы описывается дифференциальным уравнением

$$\ddot{x} = f(x) \Leftrightarrow \dot{x} = p, \;\; \dot{p} = f(x)$$

— уравнением движения материальной точки в поле потенциальных сил; здесь $f = -dU/dx$, а U — потенциал. Пусть $U(x)$ задан графиком, изображенным на рис. 1.2,*а*.

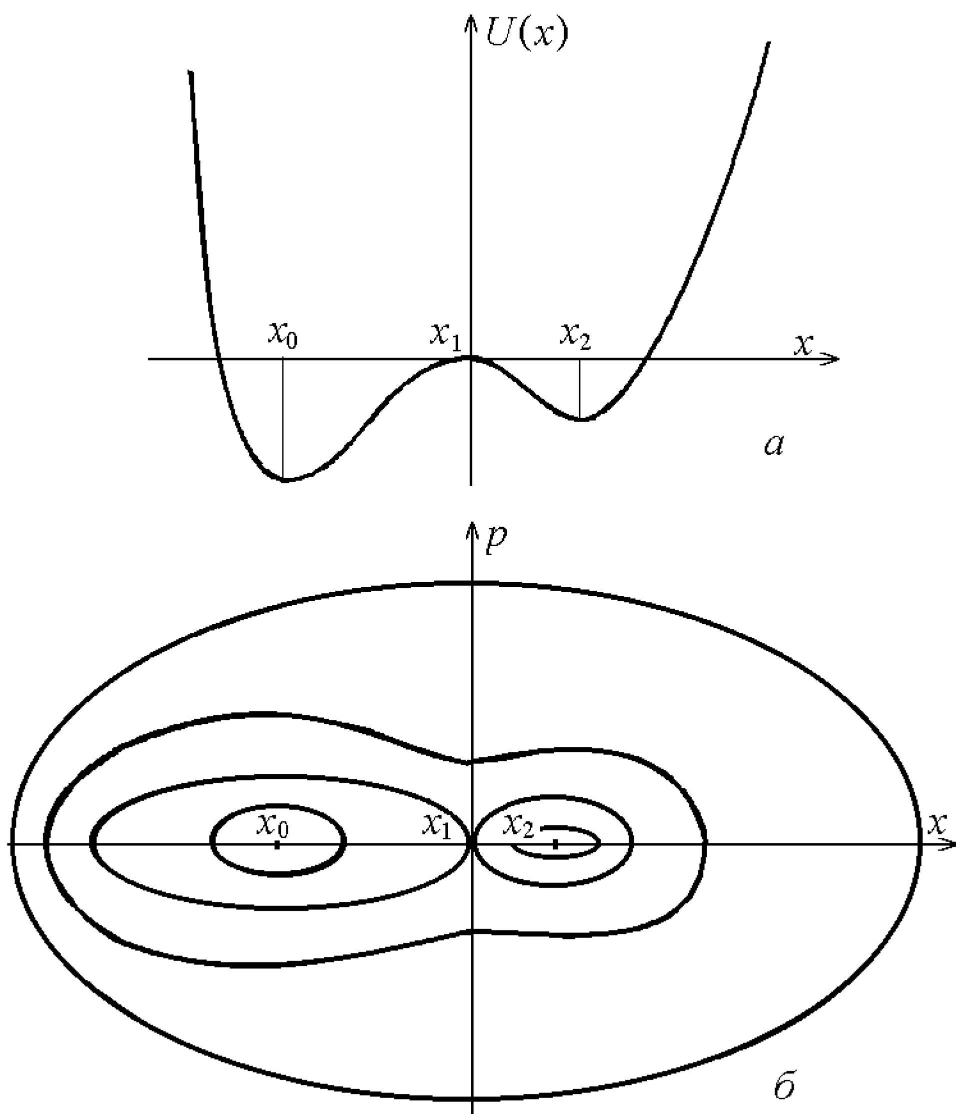


Рис. 1.2. Потенциал силового поля (*а*) и фазовые траектории (*б*) материальной точки, движущейся в потенциальном поле

Если для некоторого значения $\mathbf{x}_0$ фазовых координат системы правая часть уравнения (1.1), т.е. скорость $\mathbf{F}(\mathbf{x})$ их изменения, равна нулю, то это состояние системы не может изменяться. Точка $\mathbf{x}_0$ фазового пространства, для которой выполнено условие $\mathbf{F}(\mathbf{x}_0) = 0$, называется стационарной точкой, или положением равновесия. Для данного примера положение равновесия системы определяются условиями $p = 0$, $f(x) = 0$, т.е. стационарная точка на фазовой плоскости лежит на оси OX. Из $f(x) = 0$ следует $dU/dx = 0$,

и система имеет три положения равновесия — x_0, x_1, и x_2. В зависимости от полной энергии имеется тот или иной вид движения, изображаемый траекторией на фазовой плоскости, рис. 1.2,*б*.

Если начальное состояние таково, что его энергия не достаточна для перехода через потенциальный барьер, то точка колеблется вокруг одного из устойчивых состояний равновесия, что изображается замкнутыми фазовыми траекториями на фазовой плоскости. Если начальная энергия достаточно велика, то происходят сложные колебания точки вокруг обеих точек равновесия. Эти качественно различающиеся типы движения на фазовой плоскости разделяет кривая, называемая сепаратрисой — она проходит через точку равновесия системы. Точки равновесия системы называются особыми точками.

1.2.3. Нелинейный осциллятор. Рассмотрим динамическую систему, описываемую нелинейным эволюционным уравнением

$$\ddot{x} + \sin x = 0.$$

Аналогично случаю, рассмотренному в разделе 1.2.2, движение маятника можно рассматривать как движение в поле потенциальных сил с потенциалом $U(x) = -\cos x$, его график изображен на

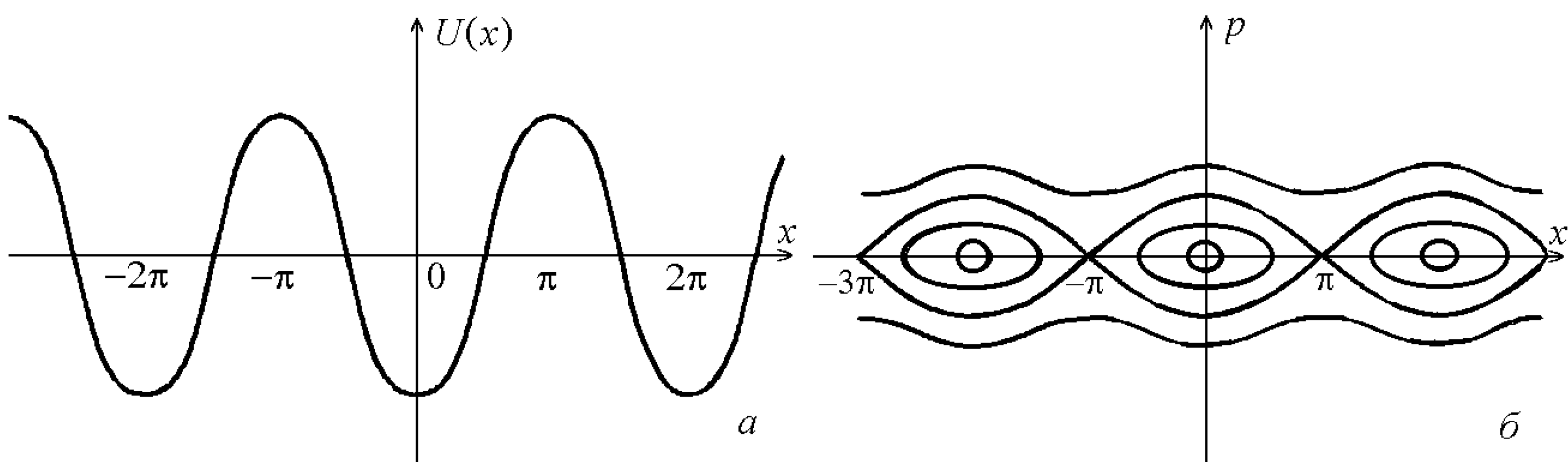


Рис. 1.3. График потенциала силового поля (*а*) и фазовые траектории (*б*) нелинейного осциллятора

рис. 1.3. Там же изображены и соответствующие фазовые траектории для различных начальных условий — фазовый портрет системы. Состояния равновесия нелинейного маятника на фазовой плоскости расположены на оси OX в точках $x_k = \pi k$, $k = 0, \pm 1, \pm 2, \ldots$, при этом $x = 0, \pm 2\pi, \pm 4\pi, \ldots$ — точки устойчивого, а $x = \pm\pi, \pm 3\pi, \ldots$ — неустойчивого равновесия. Вблизи точек устойчивого равновесия фазовые кривые представляют собой замкнутые кривые, близкие к окружностям, что выражает тот факт, что при этом маятник совершает колебания, близкие к гармоническим. Через точки неустойчивого равновесия проходит кривая, называемая сепаратрисой; она делит фазовую плоскость на области с качественно различным

поведением. При росте энергии системы колебания переходят во вращение.

1.2.4. Консервативные и диссипативные системы. В рассмотренных выше примерах полная энергия системы сохраняется. Такие системы носят название консервативных. Если же в системе возможны потери энергии, то такие системы называются диссипативными.

Рассмотрим примеры диссипативных динамических систем.

1.2.5. Маятник с затуханием. Уравнение движения

$$m\ddot{x} + \alpha\dot{x} + \beta x = 0, \quad 0 < t \leq T,$$

описывающее эволюцию системы, заменой переменных приводится к уравнению

$$\ddot{x} + 2\delta\dot{x} + x = 0, \quad 0 < \tau \leq T\sqrt{\beta/m},$$

в котором точка над символом обозначает дифференцирование по переменной τ. Здесь, в отличие от предыдущего примера, энергия не сохраняется, это — система с диссипацией. Строго говоря, это не чисто механическая система, так как здесь "сила" равна $-2\delta\dot{x} - x$ и зависит не только от положения материальной точки, но и от скорости. При $\delta = 0$ получаем рассмотренный случай консервативной системы, при $\delta \neq 0$ возникают два случая — движение периодическое и апериодическое.

Если $0 < \delta < 1$, то фазовая траектория дается решением задачи Коши, имеющим вид $x = Ae^{-\delta\tau}\cos(\omega\tau + \psi)$, $\omega = \sqrt{(1-\delta^2)}$, что соответствует затухающему колебанию, рис.1.4,*а*.

Если $\delta > 1$, то фазовая траектория описывает движение без переколебаний, рис. 1.4,*б*, и дается соотношениями $x = A_1 e^{\lambda_1\tau} + A_2 e^{\lambda_2\tau}$, $\lambda_{1,2} = (-\delta \pm \sqrt{(\delta^2 - 1)})/2$.

Итак, можно заметить, что в консервативных системах не существует притягивающих множеств, т.е. таких подмножеств фазового пространства $\mathcal{R}_N$, к которым с течением времени стремятся траектории, начинающиеся в некоторой их окрестности. Если в консервативной системе существует периодическое движение, то таких движений бесконечно много, и определяются они значением энергии при начальных условиях. Как оказалось, эти особенности являются общими для всех консервативных систем.

В диссипативных же системах могут существовать так называемые притягивающие множества, например, типа устойчивых положений равновесия, как в примере с затухающими колебаниями. Стационарные незатухающие колебания для диссипативных динамических систем нехарактерны (даже если они и существуют, то для реализации такого движения надо специальным образом выбирать начальные условия), однако оказывается, что в нелинейных системах

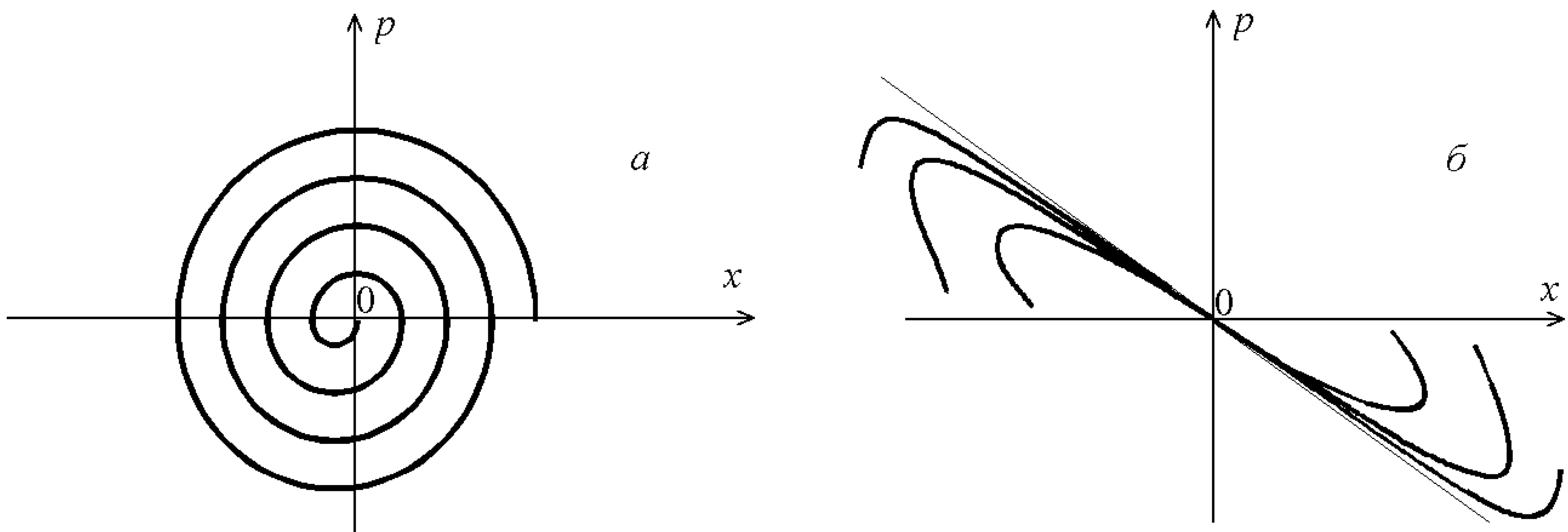


Рис. 1.4. Фазовые траектории маятника с затуханием при различных режимах демпфирования: $0<\delta<1$ (*а*); $\delta>1$ (*б*)

возможно существование периодического асимптотически устойчивого движения, так называемого автоколебания, математическим образом которого является предельный цикл Пуанкаре, изображаемый в фазовом пространстве замкнутой линией, к которой со временем стягиваются траектории из некоторой окрестности. Примером системы с предельным циклом является генератор Ван дер Поля.

1.2.6. Нелинейный осциллятор Ван дер Поля. Осциллятор Ван дер Поля описывается дифференциальным уравнением

$$\ddot{x} - a(1 - bx^2)\dot{x} + x = 0, \quad 0 < \tau \leq T,$$

здесь a — параметр возбуждения. Аналитически получить его решение не удается. На качественном уровне ясно, что при малых по модулю x коэффициент при $\dot{x}$ отрицателен — в системе возникает "отрицательное трение", отбрасывающее точку от начала координат. При больших же $|x|$ трение становится положительным, препятствующим росту координаты. Таким образом, в системе существует устойчивое состояние автоколебаний, к которому стремятся все фазовые траектории. Численное интегрирование уравнения движения осциллятора Ван дер Поля с различными начальными условиями дает фазовые траектории, изображенные на рис. 1.5.

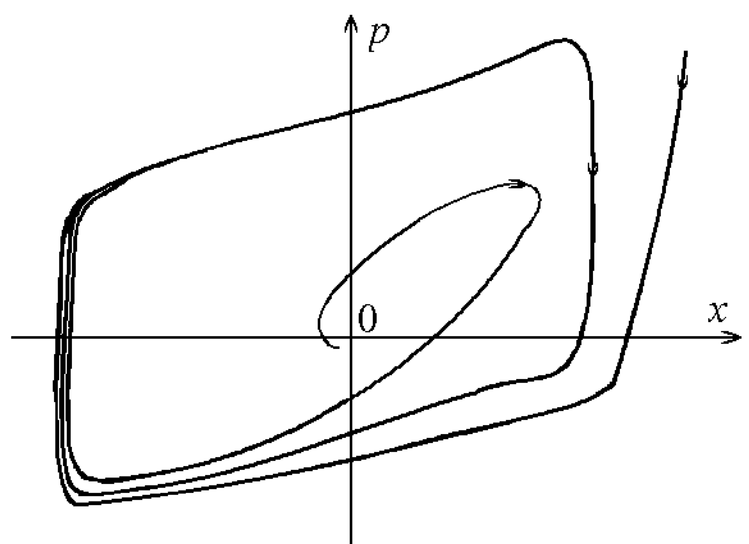


Рис. 1.5. Фазовые траектории нелинейного осциллятора Ван дер Поля

1.2.7. Странные аттракторы. При значении размерности фазового пространства $N \geq 3$ возможен принципиально новый тип фазовых траекторий — это так называемые странные аттракторы.

Фазовая траектория динамической системы в этом случае представляет собой бесконечную линию без самопересечений, причем при $t \to \infty$ траектория не покидает заданной области и не притягивается ни к точкам равновесия, ни к циклическим траекториям. Примером такой системы является аттрактор Лоренца, о котором речь пойдет ниже. Странному аттрактору соответствует сложное апериодическое движение, схожее с обыденным представлением о хаотическом процессе. Однако теоретически оно полностью предсказуемо и воспроизводимо — для начальной задачи, определяющей странный аттрактор, могут выполняться условия теоремы существования и единственности, и, задав те же начальные условия и правые части дифференциальных эволюционных уравнений, мы воспроизведем то же самое решение, и значит — ту же самую "хаотическую" траекторию. О таком поведении динамических систем говорят как о детерминированном динамическом хаосе.

1.3. Дискретные эволюционные модели

1.3.1. Разностные эволюционные уравнения. В некоторых ситуациях для моделирования динамики системы достаточно указать ее состояние в заданные дискретные моменты времени. В этом случае в качестве эволюционного оператора можно использовать функцию, выражающую состояние системы в некоторый момент времени через ее состояние в предыдущий момент. Моделью такой динамической системы может служить следующее уравнение с заданным начальным условием:

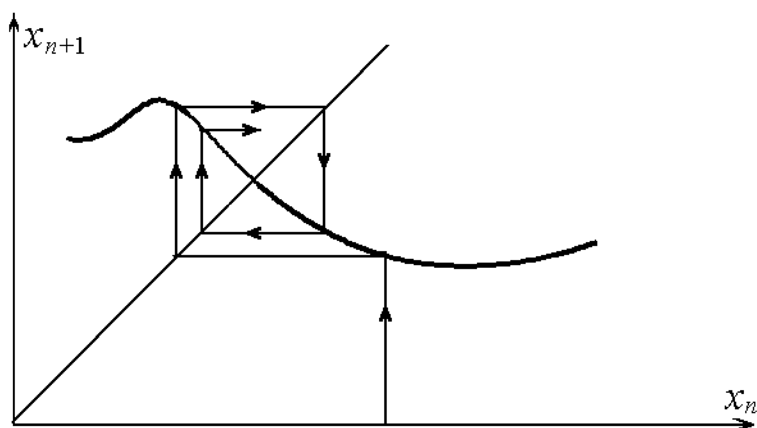


Рис. 1.6. Динамика системы с дискретным временем

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{F}(\mathbf{x}_n); \\ \mathbf{x}_0 &= \mathbf{x}_*. \end{aligned} \tag{1.4}$$

Эволюционное уравнение в (1.4) можно получить из дифференциального уравнения, приведенного в (1.1) заменой производной на разделенную разность: если, в соответствии с (1.1), $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$, то переход к разностному уравнению дает соотношение $\mathbf{x}_{n+1} = \mathbf{x}_n + \mathbf{F}(\mathbf{x}_n)\Delta t$, где Δt — шаг по времени, а $\mathbf{x}_n$ — значение параметров системы $\mathbf{x}(t)$ в момент времени $t = t_n$, $n = 1, 2, \ldots$

Если фазовое пространство одномерно, то график правой части эволюционного уравнения (1.4) — функции $F(x)$ — дает наглядное представление об эволюции системы. В частности, если этот график пересекает диагональ первой и третьей координатных четвертей,

то координаты этой точки пересечения дают стационарную точку динамической системы, так как условием пересечения графика с диагональю первого и третьего координатных углов является соотношение $x_{n+1} = x_n$. Динамику системы удобно проследить графически, как это показано на рис. 1.6.

1.3.2. Отображение Пуанкаре. Иногда для исследования системы анализ ее состояний во все моменты времени $t > 0$ является, с одной стороны, затруднительным, а с другой стороны, явно избыточным. Тогда удобно дискретное описание динамической системы, возникающее естественным образом в следующей ситуации. Рассмотрим в фазовом пространстве $\mathcal{R}_N$ поверхность $\mathcal{L}_{N-1}$ размерности $\dim \mathcal{L}_{N-1} = N-1$, на единицу меньшей, чем размерность $\dim \mathcal{R}_N = N$ фазового пространства $\mathcal{R}_N$, пусть уравнение этой поверхности задано в виде $S(\mathbf{x}) = 0$ и фазовая траектория $x(t)$, $t \geq 0$, пересекает ее под углом, отличным от нуля (т.е. не по касательной, или, как говорят, трансверсально), так, что при пересечении знак $S(\mathbf{x}(t))$ изменяется, например, с плюса на минус. Обозначим время такого первого пересечения t_1, следующего — t_2 и т.д. Уравнение $\mathbf{x}(t_k) = P(\mathbf{x}(t_{k-1}))$, связывающее координаты точек двух последующих пересечений, называется отображением Пуанкаре, а поверхность $\mathcal{L}_{N-1}$ — сечением Пуанкаре.

С помощью отображения Пуанкаре описание динамики становится проще, так как уменьшается размерность модели (точки на сечении Пуанкаре имеют размерность $N-1$), и вместо непрерывного времени рассматривается дискретный набор t_1, t_2, ... его значений. Само же описание в терминах отображения Пуанкаре в ряде случаев очень удобно. Например, стационарная точка отображения Пуанкаре, $\mathbf{x}_{k+1} = P(\mathbf{x}_k)$, соответствует циклическому движению системы по замкнутой траектории.

1.4. О решении начальных задач для дифференциальных эволюционных уравнений

1.4.1. Область определения фазовых траекторий. Пусть правые части автономной системы (1.1) удовлетворяют условию Липшица по всем своим переменным в каждой ограниченной части пространства. Тогда при любом начальном условии $\mathbf{x}(0) = \mathbf{x}_0$ существует решение $\mathbf{x}(t, \mathbf{x}_0)$ системы (1.1), определенное в некоторой окрестности начального момента времени $t = 0$. Это решение рассматривается как закон движения точки, т.е. закон изменения во времени фазовых коор-

динат точки пространства $\mathcal{R}_N$, описывающий некоторую траекторию $l = l(\mathbf{x}_0)$ в $\mathcal{R}_N$, зависящую от выбора начальной точки. Как следует из общей теории обыкновенных дифференциальных уравнений (см. например, [10]), это решение либо можно продолжить на все множество $t > 0$, либо при приближении к некоторому конечному $T > 0$ точка $\mathbf{x}(t, \mathbf{x}_0)$ уходит на бесконечность. Для простоты будем считать, что имеет место первый случай, так как второй случай сводится к первому нелинейной заменой переменной t. То же самое можно сказать и об отрицательной полуоси $t < 0$. Таким образом, будем считать, что решение $\mathbf{x}(t, \mathbf{x}_0)$ системы (1.1) задано на всей числовой оси $-\infty < t < \infty$.

1.4.2. Автономные динамические системы. Если правая часть (1.1) не зависит явно от времени, то решение начальной задачи обладает следующими свойствами:

— оно непрерывно по совокупности аргументов;

— $\mathbf{x}(0, \mathbf{x}_0) = \mathbf{x}_0$;

— $\mathbf{x}(t_2, \mathbf{x}(t_1, \mathbf{x}_0)) = \mathbf{x}(t_1 + t_2, \mathbf{x}_0)$.

Эти свойства настолько существенны, что иногда автономной системой называют систему, поведение которой удовлетворяет соотношениям 1–3. Доказательство этих свойств системы (1.1) вытекает из теоремы существования и единственности решения уравнения (1.1).

1.4.3. Типы траекторий автономных динамических систем. Решение $\mathbf{x}(t)$ автономной системы (1.1) может быть только одного из трех типов:

— непериодическое, для которого $\mathbf{x}(t_1) \neq \mathbf{x}(t_2)$ при $t_1 \neq t_2$;

— периодическое, для которого найдется такое число $T > 0$, что $\mathbf{x}(t+T) = \mathbf{x}(t)$, а $\mathbf{x}(t_1) \neq \mathbf{x}(t_2)$ при $0 \leq t_1 < t_2 < T$;

— постоянные, для которых $\mathbf{x}(t) = \mathbf{x}_0$.

Траектории, соответствующие этим трем типам траекторий, называются соответственно незамкнутой, замкнутой и точкой покоя. Замкнутая траектория иначе называется циклом.

1.4.4. Предельные точки и предельные множества. Пусть $\mathbf{x}(t)$ — некоторое решение системы (1.1) и ему соответствует некоторая траектория l в фазовом пространстве. Точка $\widehat{\mathbf{x}}$ называется предельной точкой решения, или траектории, при $t \to +\infty$, если существует последовательность моментов времени $t_k \to +\infty$, такая, что $\mathbf{x}(t_k) \to \widehat{\mathbf{x}}$. Совокупность всех таких точек называется предельным множеством при $t \to +\infty$. (Предельные точки и предельные множества встречались нам в примерах разделов 1.2.5 и 1.2.6.

Очевидно, точка покоя является своей единственной предельной точкой; замкнутая траектория также является своим собственным предельным множеством. Предельные точки незамкнутых траекторий более интересны, так как позволяют определить характер поведения траектории при больших t. Приведем свойства предельных

траекторий [10].

1. Предельное множество замкнуто (как множество в R_N).

2. Предельное множество состоит из целых траекторий, т.е. если $\widehat{\mathbf{x}} \in \widehat{l}$, то и вся траектория $l(\widehat{\mathbf{x}}) \in \widehat{l}$.

Это свойство можно сформулировать следующим образом: это множество является инвариантным относительно преобразования $\mathbf{x}_0 \to \mathbf{x}(t, \mathbf{x}_0)$, определяемом системой (1.1).

3. Для того чтобы предельное множество было пустым, необходимо и достаточно, чтобы траектория $\mathbf{x}(t)$ при $t \to \infty$ уходила в бесконечность.

4. Для того чтобы предельное множество состояло из единственной точки, необходимо и достаточно, чтобы траектория $\mathbf{x}(t)$ входила в эту точку при $t \to \infty$, т.е.

$$\mathbf{x}(t) \to \widehat{\mathbf{x}} \quad \text{при} \quad t \to \infty.$$

Глава 2

Классификация поведения динамических систем

Проведение точного количественного анализа нелинейной динамической системы возможно только в редких случаях: среди нелинейных дифференциальных уравнений, описывающих динамику систем, лишь для немногих удается получить точное решение. Однако часто на практике не столько важно знать точное численное значение параметров системы, задающее ее состояние в заданные моменты времени, сколько описание поведения системы в целом, на качественном уровне. Например, при проектировании той или иной механической конструкции важно знать, существует ли у нее состояния равновесия, какие из них устойчивы, а какие — нет, возможно ли в системе устойчивые колебательные режимы и т.п. Необходимо знать также, при каких условиях в системе возможны те или иные характерные поведения (колебания, равновесие и т.п.), и когда устойчивые состояния сменяются неустойчивыми. На эти вопросы отвечает качественная теория динамических систем.

2.1. Топологическая эквивалентность

2.1.1. Определение топологической эквивалентности. В предыдущем разделе для характеристики поведения нелинейной динамической системы использовался ее фазовый портрет. По форме фазового портрета можно судить о характерном поведении системы, причем "плавные" деформации фазового пространства, по сути, не меняют качественно динамику системы. Это соображение лежит в основе определения близости динамических систем — речь идет о так на-

зываемой топологической эквивалентности фазовых портретов.

Если фазовый портрет динамической системы, изображенный на рис. 2.1,*а* путем плавной деформации фазового пространства можно перевести в фазовый портрет, изображенной на рис.2.1,*б*, то такие системы называют топологически эквивалентными. На рис. 2.1 системы (*а*) и (*б*) топологически эквивалентны, а (*б*) и (*в*) не являются эквивалентными, например, потому, что число особых точек у этих систем различно. Эти соображения позволяют взглянуть на поведение различных динамических систем с единой точки зрения: на основе введенного понятия множество всех динамических систем можно разбить на классы, внутри которых системы демонстрируют качественно схожее поведение. Строго говоря, "плавная" деформация фазового портрета производится путем невырожденного, т.е. взаимно однозначного и взаимно непрерывного, преобразования фазовых координат, в результате которого, с одной стороны, не может появиться новых особых точек, а с другой — особые точки не могут исчезнуть, т.е. превратиться в неособые.

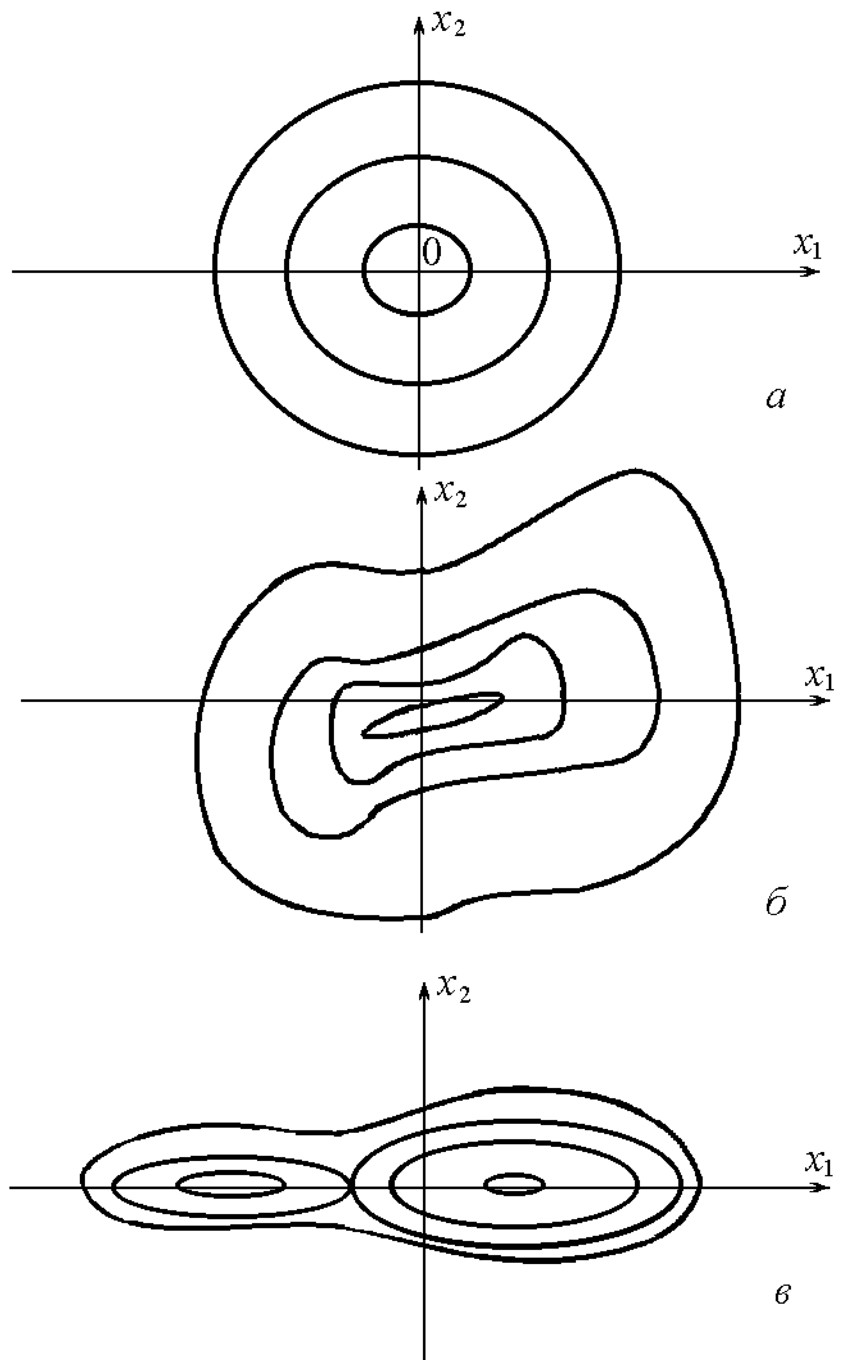


Рис. 2.1. Фазовые портреты. Системы (*а*) и (*б*) топологически эквивалентны, (*б*) и (*в*) не эквивалентны

2.1.2. Зависимость от параметра. Деформация фазового портрета может осуществляться не только заменой фазовых переменных, но и изменением параметров самой системы — например, параметра, от которого зависит потенциал системы, состоящей из материальной точки, движущейся в потенциальном силовом поле. Такие параметры носят название в н е ш н и х, или у п р а в л я ю щ и х. Казалось бы, если потенциал непрерывно зависит от параметра, то малое изменение параметра не должно приводить к качественному изменению поведения системы. Однако следующие примеры демонстрируют, что это рассуждение неверно.

Система с резонансом. Пусть первая система описывается эволю-

ционным уравнением

$$\ddot{x} - x = 0,$$

а вторая — уравнением

$$\ddot{x} - x = \varepsilon \cos t.$$

Как бы ни было мало значение ε, отличающее значение правой части второй системы от первой, поведение второй системы принципиально иное, чем поведение первой. Действительно, вторая система описывает, например, математический маятник, на который действует гармоническая внешняя сила с частотой, равной собственной частоте маятника. В таких системах возникает резонанс, при котором амплитуда колебаний линейно растет со временем. В то же время первая система при ненулевых начальных условиях демонстрирует гармонические колебания. Системы, очевидно, не являются топологически эквивалентными при любом $\varepsilon \neq 0$.

Движение в поле потенциальных сил. Рассмотрим одномерное движение материальной точки в трех ситуациях, различающихся значениями параметров, определяющих потенциал силового поля. Пусть системы описываются уравнениями

$$\ddot{x} = F_i(x) \equiv -\frac{dU_i(x)}{dx}, \quad i = 1, 2, 3,$$

причем $U_i(x) = x^4 + \varepsilon_i x^2$, $i = 1, 2, 3$, и $\varepsilon_1 > 0$, $\varepsilon_2 = 0$, $\varepsilon_3 < 0$. Графики потенциалов изображены на рис. 2.2. Особые точки здесь задаются значениями координаты x, при которых потенциал имеет локальный экстремум, так как при этом правая часть эволюционного уравнения обращается в нуль. Из выражений для потенциала и из рис. 2.2 видно,

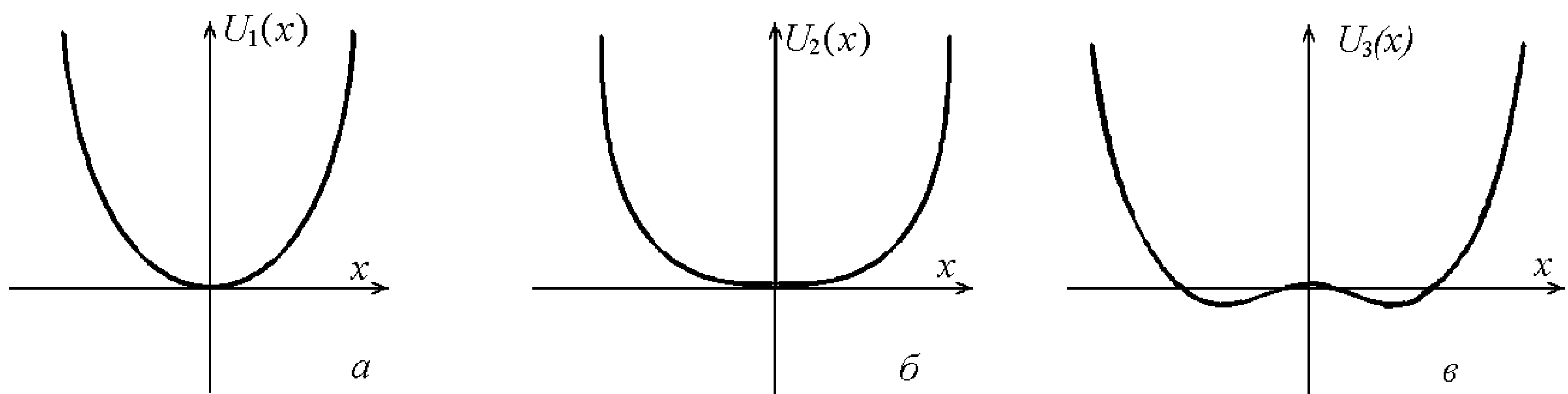


Рис. 2.2. Графики потенциалов. Системы (*а*) и (*б*) топологически эививалентны, (*б*) и (*в*) не эквивалентны

что как в первой, так и во второй системе имеется одно положение устойчивого равновесия, других же особых точек не существует. Качественно их поведение сходно — при ненулевых начальных условиях

это колебание вокруг точки равновесия. Третья же система имеет два устойчивых состояния равновесия и одно неустойчивое, спектр ее возможных поведений богаче: это и колебания вокруг одной из устойчивых положений равновесия (при малых значениях полной энергии системы), при большей же полной энергии возможно и более сложное колебание вокруг обеих точек равновесия, теоретически возможно и неустойчивое стационарное поведение (при $x = 0$). Ясно, что разумно считать первую и вторую системы топологически эквивалентны, третью же назвать им не эквивалентной.

2.2. Исследование качественного поведения систем

Следующий шаг, который мы сделаем на пути исследования нелинейных динамических систем, связан с описанием качественного поведения системы в зависимости от внешних управляющих параметров.

2.2.1. Примеры влияния управляющих параметров на динамику систем. Чем интересуются при качественном исследовании динамических систем? Чтобы ответить на этот вопрос, рассмотрим несколько примеров.

Пусть уравнение

$$\ddot{x} = F(x) = -\frac{\partial U}{\partial x},$$

описывает изменение во времени координаты материальной точки, движущейся в поле потенциальных сил; здесь U — потенциал, зависящий от параметра μ. Если потенциал имеет вид

$$U(x) = x^4 - \mu x^2,$$

(его график для $\mu < 0$ приведен на рис. 2.3,*а*, то решение описывается фазовым портретом, изображенным на рис. 2.3,*б*, с точкой устойчивого равновесия в нуле. Здесь, так же, как и в главе 1, используется обозначение $p = \dot{x}$. Если изменить знак параметра μ на противоположный, то потенциал примет вид, изображенный на рис. 2.3,*в*, при этом точка равновесия $x = 0$ становится неустойчивой, и появляются две новых устойчивых точки равновесия; соответствующий этой ситуации фазовый портрет приведен на рис. 2.3,*г*.

Если изобразить положение устойчивых точек равновесия в зависимости от параметра μ, получится диаграмма в виде вилки (рис. 2.4); отсюда произошло название "бифуркация" для точки $\mu = 0$ (на английском языке слово "fork" означает "вилка", значит, бифуркация — "двузубая вилка").

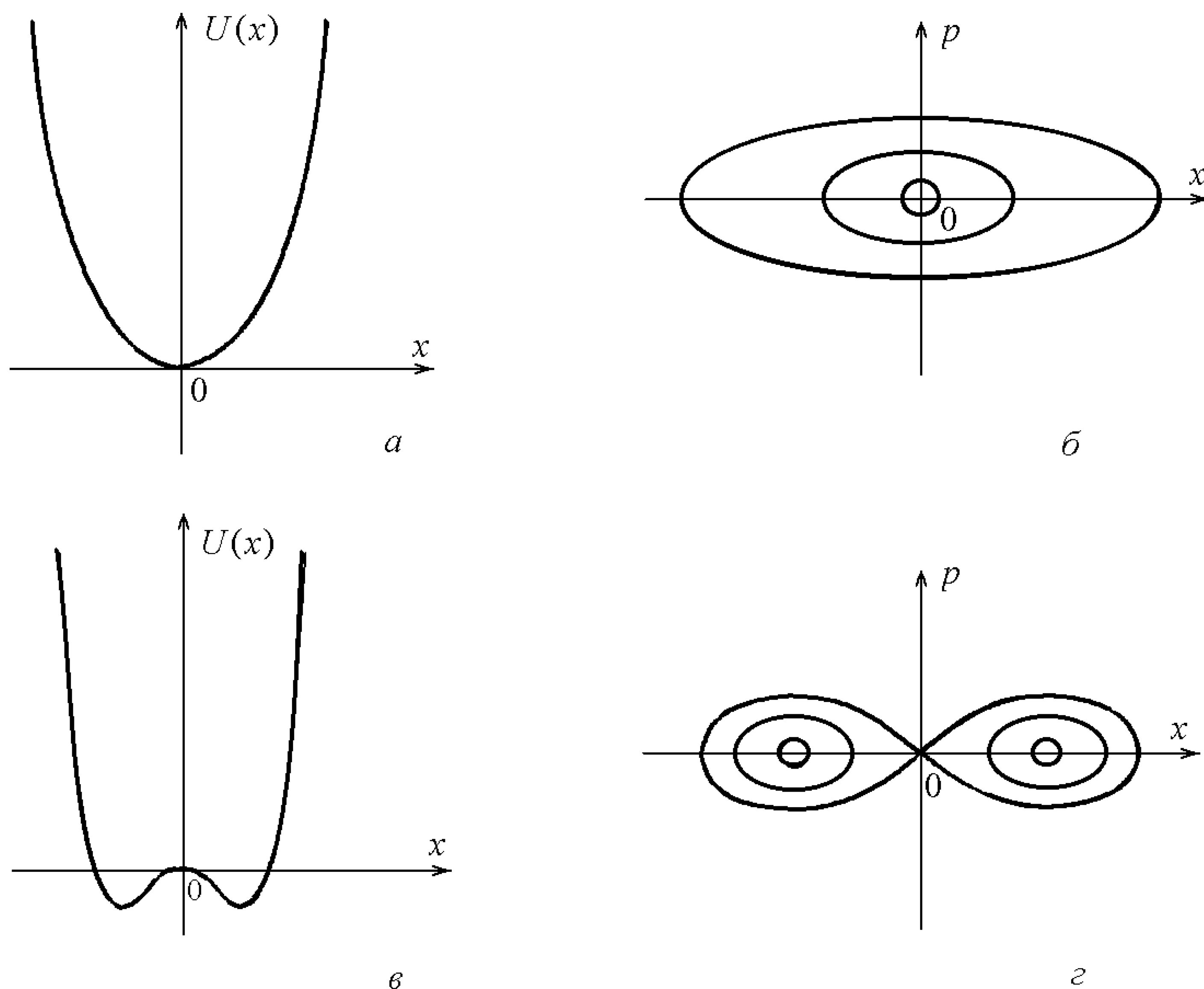


Рис. 2.3. Графики потенциала (*а*) и (*в*) и фазовые портреты (*б*) и (*г*): при $\mu < 0$ (*а*) и (*б*), при $\mu > 0$ (*в*) и (*г*)

Если изменять параметр μ от отрицательных значений к положительным, то при прохождении μ через точку бифуркации $\mu = 0$ система скачком изменяет свое поведение. При приближении значения параметра к точке бифуркации происходит выполаживание потенциала, что приводит к замедлению темпа эволюции системы; это явление носит название критического замедления и является характерным признаком грядущего качественного изменения движения системы.

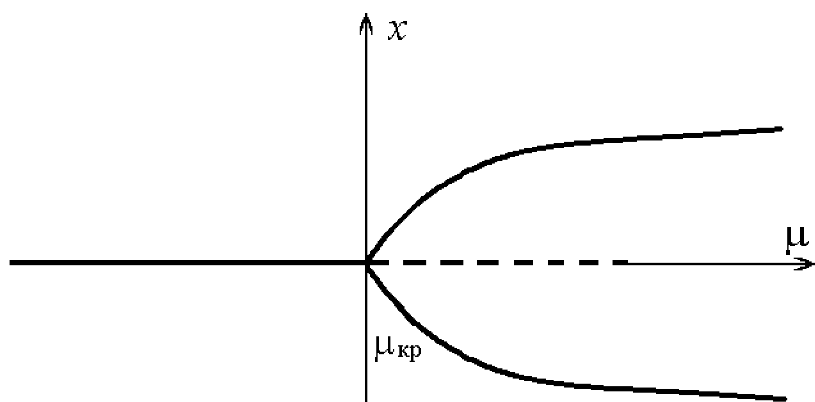


Рис. 2.4. График зависимости положения точки устойчивого равновесия от параметра (бифуркационная диаграмма). Пунктиром показано положение неустойчивого равновесия

Еще один пример связан с исчезновением и появлением предельного цикла. Рассмотрим движение шарика по желобу, имеющему вид поверхности вращения, профиль которой в зависимости от радиуса описывается формулой

$$z(r, \varphi) = U(r, \mu)$$

и изображен на рис. 2.5.

Если движение происходит достаточно медленно, так, что центробежной силой можно пренебречь по сравнению с силой тяжести, то можно заметить, что в такой системе существует движение в виде устойчивого предельного цикла — движение по окружности радиуса r_1, см. рис. 2.5,*а*; имеется также и неустойчивый предельный цикл — движение по окружности радиуса r_0.

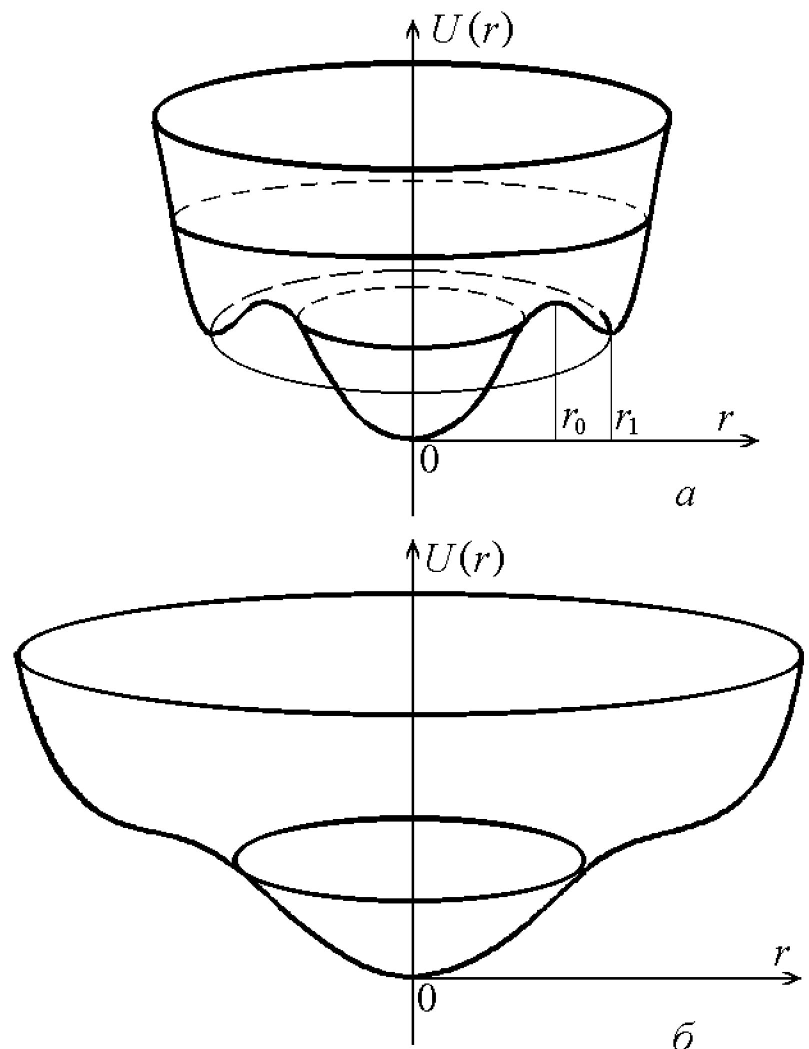


Рис. 2.5. Бифуркация исчезновения предельного цикла

Изменением значения параметра μ можно добиться выполаживания профиля (см. рис. 2.5,*б*) так, что при некотором значении μ предельные циклы исчезнут. Если μ_0 — значение параметра μ, такое, что при $\mu > \mu_0$ движение системы может происходить по предельным циклам, а при $\mu < \mu_0$ они исчезают, то такая точка μ_0 называется точкой бифуркации рождения–уничтожения предельного цикла, или точкой бифуркации Пуанкаре–Андронова–Хопфа. На рис. 2.6 изображена бифуркационная диаграмма для этого случая. Заметим, что здесь профиль желоба и график потенциала несут одну и ту же качественную информацию о поведении системы.

2.2.2. Грубые динамические системы. Обобщая примеры, рассмотренные в предыдущем разделе, заметим, что качественно поведение динамической системы можно описать, указав характерные области фазового пространства, например, области притяжения предельных множеств, окрестности седел или фокусов.

В частности, в случае двумерного фазового пространства области фазовой плоскости, в которых поведение динамической системы качественно различаются, отделяются одна от другой сепаратрисами — фазовыми траекториями, проходящими через поло-

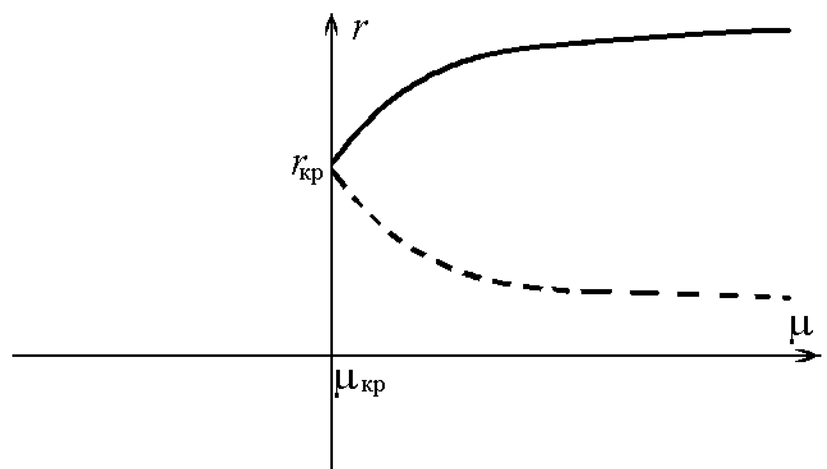


Рис. 2.6. Бифуркационная диаграмма рождения – исчезновения предельного цикла

жения равновесия (в многомерном случае, к сожалению, сепаратрисы не имеют столь наглядной интерпретации, так как при $\dim \mathcal{R}_N > 2$ фазовое пространство, вообще говоря, не разбивается на области, заполненные траекториями с одинаковыми предельными множествами). С примером сепаратрисы на плоскости мы уже встречались, когда рассматривали физический маятник; в этом примере она отделяет область фазового пространства, соответствующую колебательному движению, от области, соответствующей вращению (рис. 2.7).

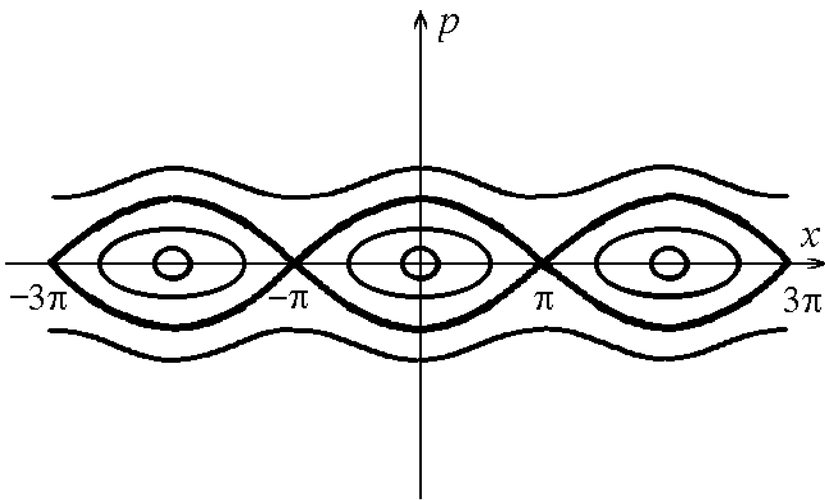


Рис. 2.7. Области колебания и вращения физического маятника на фазовой плоскости разделяются сепаратрисой

Значительный интерес в теории динамических систем представляет вопрос о поведении системы при изменении параметров эволюционного оператора. Если эти параметры не являются функциями состояния системы, их называют внешними, или управляющими параметрами. Поведение системы при фиксированных внешних управляющих параметрах в значительной степени определяется типом ее особых точек — стационарных точек равновесия. Рассмотрим, например, материальную точку, движущуюся в вязкой среде, в которой сила трения, пропорциональная скорости движения $\dot{x}$, настолько велика, что не позволяет точке двигаться с ускорением, так, что скорость движения оказывается пропорциональной действующей внешней силе. Эволюционное уравнение для такой системы запишется в виде

$$\dot{x} = F(x, \mu) = -\frac{\partial U(x, \mu)}{\partial x},$$

где μ — управляющий параметр, x — скалярная координата материальной точки, $F(\cdot, \cdot)$ — потенциальная сила с потенциалом $U(\cdot, \cdot)$ с изолированными минимумами в точках $x = x_1$ и $x = x_2$, соответствующими устойчивым состояниям равновесия системы. В силу свойств непрерывных функций (а $U(x)$ по смыслу даже дифференцируема) обязательно найдется точка x_3 на интервале (x_1, x_2), в которой система может находится в состоянии неустойчивого равновесия — это точка максимума функции $U(x)$. Если других особых точек в системе нет, то заданием точек x_1, x_2 и x_3 полностью исчерпано качественное описание поведения динамической системы при фиксированном управляющем параметре μ, так как при $x \in (-\infty, x_1) \cup (x_2, x_3)$ потенциал $U(x)$ убывает, что соответствует ситуации, когда на точку действует сила, заставляющая ее двигаться в направлении

слева направо, а при $x \in (x_1, x_2) \cup (x_3, \infty)$ потенциал $U(x)$ возрастает, т.е. сила заставляет точку двигаться в направлении справа налево.

В аналогичном примере системы с двумя фазовыми координатами (x_1, x_2), если потенциал $U(x_1, x_2)$ имеет два изолированных минимума, то возможно существование некоторого пути на плоскости (x_1, x_2), соединяющего точки минимума и проходящего через седло функции $U(x_1, x_2)$ — неустойчивую точку равновесия. Если нет других особых точек, то плоскость (x_1, x_2) делится на два "бассейна" притяжения к точкам минимума: система, находящаяся в бассейне притяжения одного из точек минимума, имеет тенденцию к движению в направлении градиента U к этой точке минимума. Если при изменении управляющего параметра μ происходит лишь изменение положений особых точек, но особые точки не исчезают и новые не возникают, то с качественной точки зрения поведение системы при разных значениях параметра — одно и то же; иными словами, фазовые портреты систем оказываются топологически эквивалентными. Как уже указывалось, качественное изменение фазового портрета системы в зависимости от внешнего параметра и называется бифуркацией, а значение параметра μ, при котором это происходит, называется бифуркационным. Бифуркации удобно исследовать путем анализа изменений фазовых траекторий в расширенном комбинированном пространстве, представляющем прямое произведение фазового пространства на пространство параметров.

Сохранение структуры фазового пространства, определяемого существованием притягивающих множеств, состояний равновесия и других особых точек динамической системы, связано с понятиями грубости, структурной устойчивости и топологической эквивалентности.

Грубыми, или структурно устойчивыми, называются такие динамические системы, которые при малых возмущениях качественно не меняют структуры разбиения пространства параметров на области, отвечающие различным типам решений. Более аккуратно, для динамических систем, описывающихся дифференциальными уравнениями, структурно устойчивыми называются такие системы, которые малые дифференцируемые возмущения эволюционного оператора приводят к топологически эквивалентным динамическим системам, т.е. таким, для которых возмущенный фазовый поток можно привести в невозмущенный путем невырожденной замены фазовых координат.

Рассмотрим возмущения динамической системы, описываемые путем изменения параметра оператора эволюции. Изменение параметра, как уже говорилось, может оставлять систему качественно в том же состоянии, либо привести к принципиальному изменению характера эволюции системы (как в примере с возникновением автоколебаний в генераторе Ван-дер-Поля при изменении параметра и превышении им

некоторого значения, которое называется порогом генерации). Смена одних устойчивых состояний другими при изменении управляющих параметров вызывает последовательность фазовых переходов динамической системы от одних грубых (структурно устойчивых) режимов к другим через негрубое состояние в точке бифуркации.

Важный аспект качественной теории нелинейных динамических систем — исследование и классификация точек равновесия и нахождение предельных циклов.

2.2.3. Классификация особых точек. Займемся сначала изучением точек равновесия, для чего рассмотрим автономную систему с двумерным фазовым пространством, описываемую автономными эволюционными уравнениями

$$\begin{aligned} \dot{x}_1 &= F_1(x_1, x_2), \\ \dot{x}_2 &= F_2(x_1, x_2), \end{aligned} \tag{2.1}$$

дополненными некоторыми начальными условиями. Поделим первое уравнение системы (2.1) на второе, и получим

$$\frac{dx_1}{dx_2} = \frac{F_1(x_1, x_2)}{F_2(x_1, x_2)}; \tag{2.2}$$

в тех точках, где знаменатель в правой части (2.2) обращается в нуль, поменяем делимое и делитель местами и получим уравнение

$$\frac{dx_2}{dx_1} = \frac{F_2(x_1, x_2)}{F_1(x_1, x_2)}. \tag{2.3}$$

Для уравнений (2.2) и (2.3) рассмотрим понятие особых точек. Пусть уравнение (2.2) или (2.3) рассматривается в области G, точка P принадлежит области G или ее границе (где уравнение (2.2) или (2.3) тоже может быть задано). Если существует окрестность точки P (т.е. круг достаточно малого радиуса с центром в точке P), такая, что через каждую точку этой окрестности проходит одна и только одна интегральная линия, и хотя бы одна из функций $F_1(x_1, x_2)/F_2(x_1, x_2)$, $F_2(x_1, x_2)/F_1(x_1, x_2)$, непрерывна в этой точке, то P — неособая точка.

Если P не является неособой, то она называется особой точкой. При этом возможны три случая.

1. Точка P принадлежит границе области G.

2. Точка P — точка, в которой существует не одно решение уравнений (2.2) или (2.3), т.е. в любой ее окрестности найдется более одной интегральной линии, либо точка P — предельная точка множества точек неединственности.

3. Поле направлений в точке P имеет разрыв.

Возможны и комбинации этих трех случаев, например, граничная точка неединственности и пр.

Особый интерес представляют изолированные особые точки, т.е. те, в окрестности которых не имеется других особых точек.

Если $F_i(x_1, x_2)$, $i = 1, 2$, имеют непрерывные частные производные по x_1 и x_2, то все внутренние точки области, где рассматривается уравнение (2.2), (2.3), в которых $F_1(x_1, x_2)$ и $F_2(x_1, x_2)$ не обращаются одновременно в нуль, не являются особыми.

Рассмотрим теперь точку $(x_{1,0}, x_{2,0})$ на фазовой плоскости, в которой выполнены равенства $F_1(x_{1,0}, x_{2,0}) = 0$ и $F_2(x_{1,0}, x_{2,0}) = 0$ (не ограничивая общности, можно считать $x_{1,0} = x_{2,0} = 0$). Представим функции $F_i(\cdot, \cdot)$, $i = 1, 2$, по формуле Тейлора в окрестности точки $(x_{1,0}, x_{2,0}) = (0, 0)$, и подставим в (2.2):

$$\frac{dx_1}{dx_2} = \frac{F'_{1x_1}(0,0)x_1 + F'_{1x_2}(0,0)x_2 + o(x_1^2 + x_2^2)}{F'_{2x_1}(0,0)x_1 + F'_{2x_2}(0,0)x_2 + o(x_1^2 + x_2^2)}. \qquad (2.4)$$

Это соотношение не определяет значение производной dx_1/dx_2 в точке $(0, 0)$, но если

$$\det \begin{pmatrix} F'_{1x_1}(0,0) & F'_{1x_2}(0,0) \\ F'_{2x_1}(0,0) & F'_{2x_2}(0,0) \end{pmatrix} \neq 0,$$

то как бы ни была доопределена производная dx_1/dx_2 в точке $(0, 0)$, эта точка будет точкой разрыва поля направлений, и значит — особой точкой уравнения (2.2).

2.2.4. Поведение вблизи особых точек. Исследуем теперь поведение интегральных линий уравнения (2.2), (2.3) в окрестностях особых точек. Если окрестность достаточно мала, то на характер их поведения влияют лишь главные, линейные члены, поэтому можно рассматривать "линеаризованное" уравнение

$$\frac{dx_1}{dx_2} = \frac{ax_1 + bx_2}{cx_1 + dx_2},$$

полагая при этом, что $ad - bc \neq 0$. Линейным неособым преобразованием

$$\eta = \eta(x_1, x_2),$$
$$\xi = \xi(x_1, x_2)$$

эти уравнения можно привести к одному из трех канонических уравнений: либо к уравнению вида

$$\frac{d\eta}{d\xi} = k\frac{\eta}{\xi}, \quad k \neq 0, \qquad (2.5)$$

либо — к уравнению вида

$$\frac{d\eta}{d\xi} = \frac{\xi + \eta}{\xi}, \tag{2.6}$$

либо, наконец, к уравнению вида

$$\frac{d\eta}{d\xi} = \frac{\xi + k\eta}{k\xi - \eta}. \tag{2.7}$$

В первом случае общим интегралом служит уравнение

$$a\eta + |b\xi|^k = 0.$$

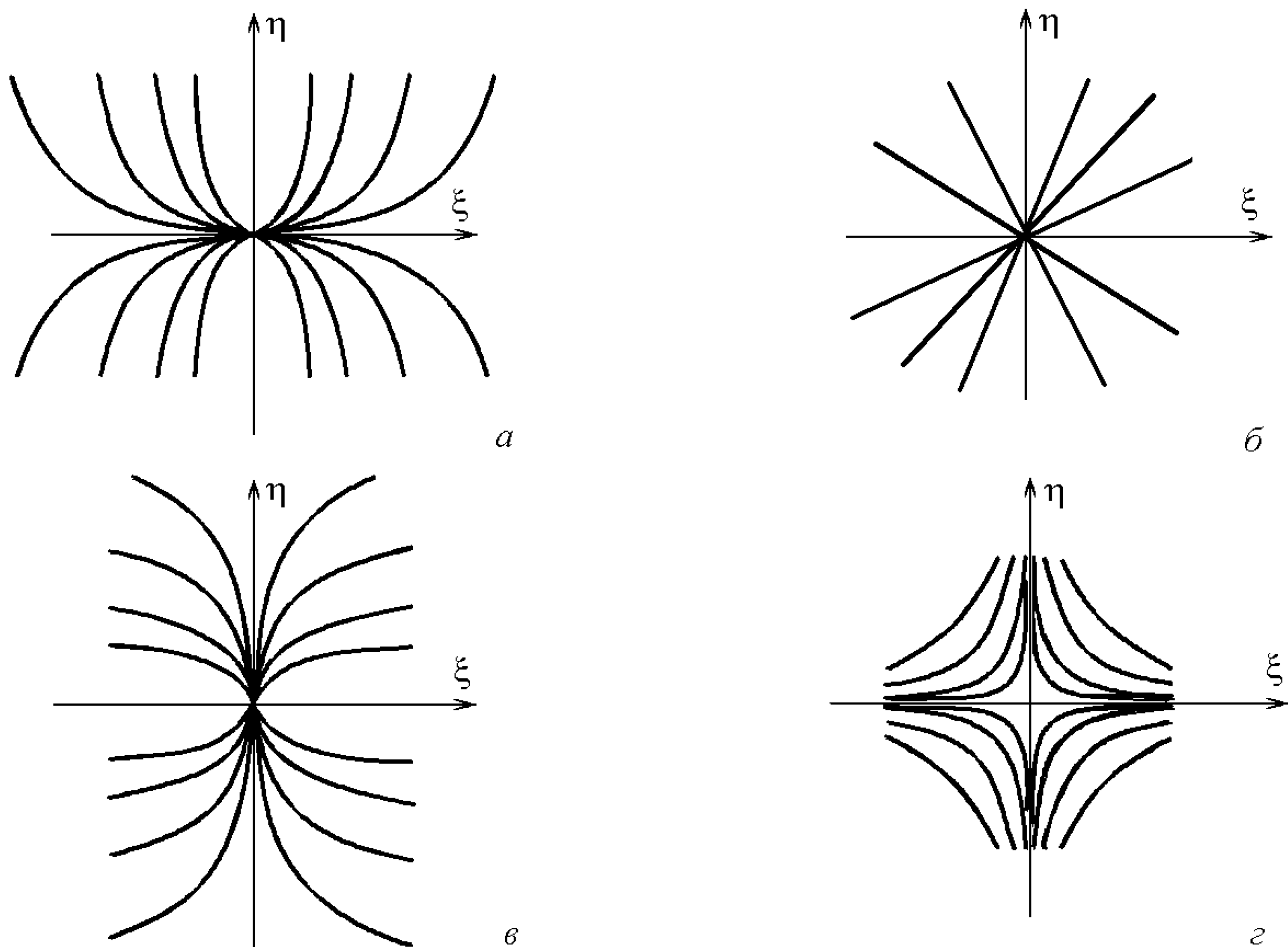


Рис. 2.8. Интегральные линии вблизи особых точек: (*а*), (*б*) и (*в*) – узлы, (*г*) – седло

Поведение интегральных линий при $k > 1$ схематически изображено на рис. 2.8,*а*; здесь все интегральные линии касаются в точке O оси $O\xi$ (за исключением обеих половин оси $O\eta$). Сами оси $O\xi$ и $O\eta$ являются интегральными линиями уравнения (2.5), за исключением точки O.

Случай $k = 1$ соответствует интегральным линиям, изображенным на рис. 2.8,*б*.

При $0 < k < 1$ все интегральные линии касаются оси $O\eta$ (за исключением обеих половин оси $O\xi$), рис. 2.8,*в*.

Итак, во всех случаях, когда $k > 0$, всякая интегральная линия подходит к точке O по определенному направлению. Вообще, когда всякая интегральная линия, имеющая точки, достаточно близкие к O, приближаются к O как угодно близко и по определенному направлению, то говорят, что точка O является *узлом*. Поэтому в рассматриваемом случае при $k > 0$ точка O есть узел для интегральных линий уравнения (2.5).

Поведение интегральных линий уравнения (2.5) при $k < 0$ изображено на рис. 2.8,*г*. В этом случае к точке O сколь угодно близко подходят лишь четыре интегральных линии — две полуоси $O\eta$ и две полуоси $O\xi$, всякая же другая интегральная линия, приблизившись достаточно близко к точке O, начинает от нее удаляться. Такие особые точки называются седлами.

Рассмотрим теперь интегральные линии уравнения (2.6). Его общим интегралом служит уравнение

$$b\eta = \xi(a + b \ln |\xi|).$$

Все интегральные линии касаются в точке O оси $O\eta$, рис. 2.9,*а*, из координатных осей только $O\eta$ является интегральной линией. Точка O и в этом случае является узлом.

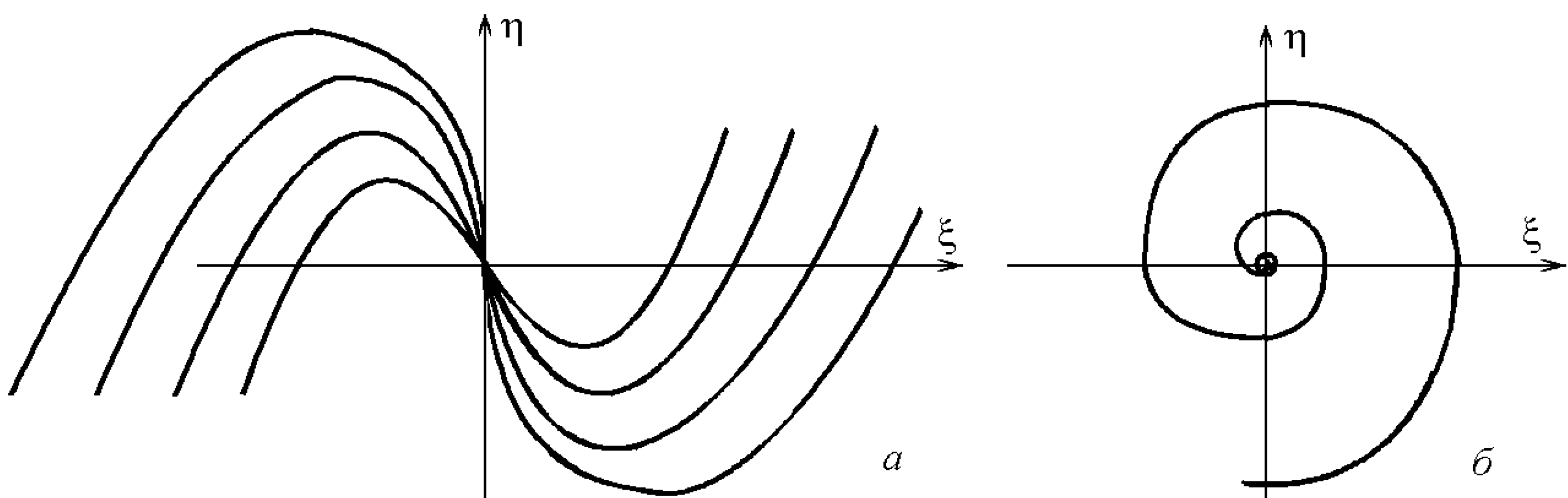


Рис. 2.9. Интегральные линии вблизи особой точки: узел (*а*), фокус (*б*)

Наконец, рассмотрим уравнение (2.7). Оно легко интегрируется при переходе к полярным координатам. Положим

$$\xi = \rho \cos \varphi, \qquad \eta = \rho \sin \varphi,$$

после такой замены переменных получим вместо (2.7) уравнение

$$\frac{d\rho}{d\varphi} = k\rho,$$

решением которого является

$$\rho = Ce^{k\varphi}.$$

Если $k > 0$, то все интегральные линии приближаются к O, бесконечно навиваясь на эту точку при $\varphi \to -\infty$; если же $k < 0$, то то же самое происходит при $\varphi \to -\infty$, см. рис. 2.9,*б*. В этом случае нельзя указать определенного направления, по которому интегральная линия приближается к O; такие особые точки носят название ф о к у с о в.

При $k = 0$ семейство интегральных линий состоит из концентрических окружностей с центром в точке O. При этом точка O называется ц е н т р о м. Вообще центром называется такая особая точка, для которой существует окрестность, целиком заполненная замкнутыми интегральными линиями, содержащими внутри себя эту точку. Заметим, однако, что центр может перейти в фокус, если к числителю и знаменателю линеаризованного уравнения приписать члены сколь угодно малого порядка, поэтому в случае особой точки в виде центра поведение интегральных линий исходного уравнения (2.2), (2.3) не определяется членами только первого порядка.

Изложенная классификация особых точек принадлежит Пуанкаре.

2.3. Устойчивость динамических систем

2.3.1. Устойчивость особых точек. Вернемся к системе уравнений (2.1), описывающей эволюцию изучаемой автономной динамической системы. Необходимым и достаточным условием того, чтобы точка $\mathbf{x}_0 = (x_{1,0}, x_{2,0})$ была точкой равновесия, является равенство нулю правой части системы (1): $F_i(x_{1,0}, x_{2,0}) = 0$, $i = 1, 2$. Это означает, что точка покоя (равновесия) является особой точкой уравнения (2.2), (2.3). Исследуем устойчивость точек покоя динамической системы (2.1), для чего в окрестности особой точки запишем линеаризованную систему уравнений, полученную из (2.1) отбрасыванием членов более высокого порядка, чем линейные:

$$\begin{aligned} \dot{x}_1 &= ax_1 + bx_2, \\ \dot{x}_2 &= cx_1 + dx_2. \end{aligned} \tag{2.8}$$

Заменим переменные $\xi = \xi(x_1, x_2)$, $\eta = \eta(x_1, x_2)$ и приведем уравне-

ния (2.8) к виду

$$\dot{\xi} = \lambda_1 \xi,$$
$$\dot{\eta} = C\xi + \lambda_1 \eta,$$

либо к виду

$$\dot{\xi} = \lambda_1 \xi,$$
$$\dot{\eta} = \lambda_2 \eta,$$

если собственные числа матрицы линеаризованной системы не кратны. Рассмотрим сначала второй случай. Если $Re\lambda_1 > 0$, $Re\lambda_2 > 0$, то легко видеть, что точка равновесия представляет собой неустойчивый фокус. Если $Re\lambda_1 < 0$ и $Re\lambda_2 < 0$, то особая точка может быть либо устойчивым узлом, либо устойчивым фокусом. Если знаки корней различны, то особая точка является седлом.

Случай с кратным корнем характеристического уравнения не приводит к качественным изменениям решения, так как решение системы дается формулами $\xi = C_1 e^{\lambda_1 t}$, $\eta = C_2 t e^{\lambda_1 t}$.

2.3.2. Предельные циклы. Перейдем теперь к вопросам анализа динамических систем, демонстрирующих поведение в виде предельного цикла. Решить вопрос о существовании предельного цикла в системе с двумя фазовыми координатами помогает теорема Пуанкаре–Бендиксена.

Выберем неособую точку $\mathbf{x}^{(0)} = (x_1^0, x_2^0)$ фазового пространства $\mathcal{R}_2$ и возьмем ее в качестве начального значения системы (1). Траектория системы при $t > 0$ на фазовой плоскости изображается в виде кривой с началом в $\mathbf{x}^{(0)}$ — она называется полутраекторией.

Теорема Пуанкаре–Бендиксена. *Пусть число координат системы $N = 2$, и полутраектория при $t > 0$ остается в ограниченной области, не приближаясь к особой точке, то эта траектория — либо предельный цикл, либо приближается к предельному циклу.*

Поясним, как можно использовать теорему Пуанкаре–Бендиксена. Пусть на плоскости задана область D, ограниченная внешней и внутренней кривой, такая, что:

— все траектории входят в D;

— в $\overline{D}$ нет особых точек.

Тогда условия теоремы выполнены.

В многомерном случае все несколько сложнее, так как помимо узлов, фокусов, седел, центров и предельных циклов многомерные динамические системы могут демонстрировать и значительно более сложное поведение. Речь идет о так называемых странных аттракторах; они рассматриваются ниже, в разделах, посвященных динамическому хаосу.

2.3.3. Устойчивость по Ляпунову. Рассмотрим теперь понятие устойчивости динамических систем по Ляпунову. Пусть

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}), \quad \mathbf{x} \in \mathcal{R}_N, \quad \mathbf{F} \in \mathcal{R}_N. \tag{2.9}$$

$$\mathbf{x}(0) = \mathbf{x}_0 \tag{2.10}$$

и $\overline{\mathbf{x}}(t)$ — фазовая траектория. Движение называется устойчивым по Ляпунову, если для любого $\varepsilon > 0$ найдется такое число $\delta = \delta(\varepsilon)$, что для любой фазовой траектории $\widetilde{\mathbf{x}}(t)$, такой, что $\|\overline{\mathbf{x}}(0) - \widetilde{\mathbf{x}}(0)\| < \delta$, выполняется неравенство $\|\overline{\mathbf{x}}(t) - \widetilde{\mathbf{x}}(t)\| < \varepsilon$ для всех $t > 0$; иными словами, устойчивость по Ляпунову означает, что все другие траектории, находившиеся рядом с $\mathbf{x}_0$ в начальный момент времени, при $t > 0$ не будут далеко уклоняться от $\overline{\mathbf{x}}(t)$. Если траектории сближаются при $t \to \infty$, то говорят об асимптотической устойчивости.

2.3.4. Орбитальная устойчивость. Устойчивость по Ляпунову характеризует близость состояний системы в один и тот же момент времени. Однако можно говорить и об устойчивости качественного поведения динамической системы. Один из подходов, связанный

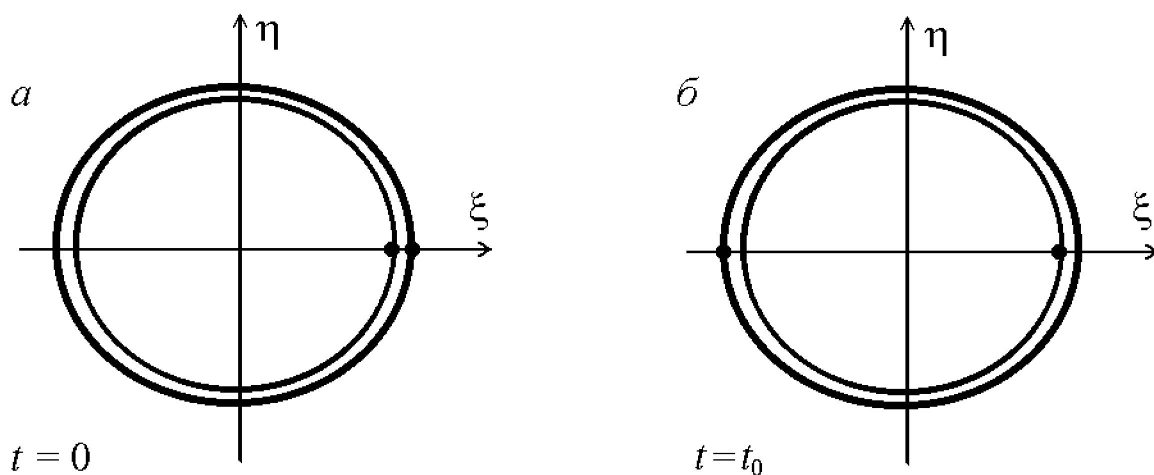


Рис. 2.10. При орбитальной устойчивости контролируется близость фазовых траекторий: при $t = 0$ состония систем близки (*а*); при $t = t_0$ состояния систем сильно различаются (*б*)

с понятием топологической эквивалентности динамических систем, рассматривался нами в разделе 2.1. Этим же целям служит понятие о р б и т а л ь н о й у с т о й ч и в о с т и, которая характеризует не поведение системы в заданный момент времени, а структуру всей фазовой траектории в целом. Понятие орбитальной устойчивости дается на качественном уровне: если некоторая фазовая траектория находится в достаточной близости от траектории $\overline{\mathbf{x}}(t)$, то эта близость сохраняется с течением времени, однако при орбитальной устойчивости точки, находящиеся на разных траекториях и близкие в некоторый момент времени, с течением времени могут сколь угодно далеко разойтись друг от друга.

Понятие орбитальной устойчивости можно проиллюстрировать

следующим примером. Рассмотрим динамическую систему, заданную эволюционными уравнениями

$$\dot{\varphi} = r,$$
$$\dot{r} = 0.$$

Семейство ее фазовых траекторий состоит из концентрических окружностей с центром в точке O, причем на разных окружностях движение происходит с разной угловой скоростью. Траектории, близкие в начальный момент времени, остаются близкими при $t > 0$, однако координата φ на разных траекториях в заданный момент времени могут различаться на полпериода (рис. 2.10).

2.3.5. Устойчивость и ляпуновские характеристические показатели. Вернемся к устойчивости по Ляпунову. Пусть $\mathbf{x}^{(0)}(t)$ — решение системы (2.9), (2.10), которое требуется исследовать на устойчивость. Запишем $\mathbf{y} = \mathbf{x}(t) - \mathbf{x}^{(0)}(t)$, и подставив в (2.9), получим

$$\dot{\mathbf{y}} = A(t)\mathbf{y}, \tag{2.11}$$

где $A(t) = A_{ij}(t) = \partial F_i/\partial x_j$ в точке $\mathbf{x}^{(0)}(t)$. Уравнение (2.11) называется уравнением в вариациях. Пусть $\mathbf{Y}(t)$ — фундаментальная система решений, так, что

$$\dot{\mathbf{Y}} = A(t)\mathbf{Y},$$

тогда любое решение (2.11) можно записать в виде

$$\mathbf{y}(t) = \mathbf{Y}(t)\mathbf{y}(t_0), \tag{2.12}$$

Линейная система (2.11) устойчива по Ляпунову тогда и только тогда, когда любое решение (2.12) ограничено, следовательно, для устойчивости движения $\mathbf{Y}$ должна быть ограниченной.

Устойчивость системы (2.9), (2.10) можно охарактеризовать в терминах показателей Ляпунова.

Определение. *Характеристическим показателем Ляпунова* функции $\Phi(t)$ называется действительное число

$$L[\Phi(t)] = \limsup_{t\to\infty}(t^{-1}\ln \mid \Phi(t) \mid).$$

Для экспоненты $\Phi(t) = e^{\alpha t}$ соответствующий характеристический показатель Ляпунова $L[\Phi(t)]$ равен ее показателю α, это означает, что $L[\Phi(t)]$ дает характеристику степени роста функции $\Phi(t)$ в сравнении с экспонентой.

Пусть $\mathbf{y}^i(t)$ — i-е фундаментальное решение системы (2.11). Тогда характеристические показатели системы (2.9) определяются соотношениями

$$\lambda_i = \limsup_{t\to\infty}(t^{-1}\ln\|\mathbf{y}^i(t)\|), \quad i = 1, \ldots, N.$$

Так как матрица A ограничена, то и λ_i конечны. Числа λ_i, $i = 1, \ldots, N$, называются обобщенными характеристическими показателями траектории $\mathbf{x}^{(0)}(t)$, а если они упорядочены по невозрастанию — то спектром ляпуновских характеристических показателей траектории $\mathbf{x}^{(0)}(t)$. Знаки ляпуновских характеристических показателей, расположенных по невозрастанию, называются сигнатурой спектра.

Если $\mathbf{x}^{(0)}(t)$ — состояние равновесия, то $\mathbf{F}(\mathbf{x}^{(0)}(t)) = 0$, $A(t)$ не зависит от времени и решение линеаризованного уравнения в окрестности точки $\mathbf{x}^{(0)}(t)$ записывается в виде

$$\mathbf{y}(t) = e^{(At)}\mathbf{y}(t_0)$$

и поэтому решение устойчиво по Ляпунову, если все собственные числа s_i матрицы A имеют отрицательные действительные части. В этом случае, согласно определению, $\lambda_i = Re\, s_i$, $i = 1, \ldots, N$, и условием асимптотической устойчивости по Ляпунову является сигнатура “$-, -, -, -, \ldots, -$”. Обращение в нуль старшего ляпуновского показателя требует особого анализа.

2.3.6. Устойчивость периодических решений. Рассмотрим некоторое периодическое решение $\mathbf{x}^{(0)}(t)$ системы (2.9), и пусть число $T > 0$ является периодом, т.е. $\mathbf{x}^{(0)}(t) = \mathbf{x}^{(0)}(t + T)$, тогда в силу определения матрица A линеаризованной системы (2.11) тоже периодична с периодом T. В самом деле, в силу автономности исходной системы (2.9) ее правая часть $\mathbf{F}$ от времени не зависит явно, аргументом ее является функция $\mathbf{x}^{(0)}(t)$, периодическая по t с периодом T. Матричные элементы матрицы A являются производными от $\mathbf{F}$ по x, вычисленными в заданной “точке” $\mathbf{x}^{(0)}(t)$; они также не зависит от времени явно, но их неявная зависимость от t остается, она определяется зависимостью $\mathbf{x}^{(0)}(t)$ от времени, и значит, тоже является периодической. Следовательно,

$$\dot{\mathbf{y}}(t+T) = A(t+T)\mathbf{y}(t+T) = A(t)\mathbf{y}(t+T),$$

а значит если $\mathbf{Y}(t)$ — матрица фундаментальных решений линеаризованной системы (2.11), то и $\mathbf{Y}(t + T)$ тоже матрица ее фундаментальных решений, причем $\mathbf{Y}(t + T) = \mathbf{Y}(t)\mathbf{Y}(T)$. Для доказательства последнего утверждения заметим, что для нахождения решения начальной задачи для линеаризованного уравнения в точке $t + T$ с любым начальным условием $\mathbf{y}(0) = \mathbf{y}_0$ можно воспользоваться стандартной формулой $\mathbf{y}(t + T) = \mathbf{Y}(t + T)\mathbf{y}_0$. С другой стороны, вычислив сначала решение $\mathbf{y}(T) = \mathbf{Y}(T)\mathbf{y}_0$ в момент времени T, пользуясь периодичностью матрицы A, найдем $\mathbf{y}(t + T)$ как решение начальной задачи

$$\dot{\mathbf{y}}(t) = A(t)\mathbf{y}(t), \quad t \geq T, \quad \mathbf{y}(T) = \mathbf{Y}(T)\mathbf{y}_0,$$

в виде $\mathbf{y}(t+T) = \mathbf{Y}(t)\mathbf{Y}(T)\mathbf{y}_0$. Сравнивая оба выражения для $\mathbf{y}(t+T)$, в силу произвольности начального условия $\mathbf{y}_0$ получим требуемое равенство $\mathbf{Y}(t+T) = \mathbf{Y}(t)\mathbf{Y}(T)$.

Матрица (оператор) $\mathbf{Y}(T)$ называется матрицей (оператором) монодромии. Заметим, что эта матрица не зависит от t. Ее собственные значения ρ_i, $i = 1, \ldots, N$, называются мультипликаторами периодического решения $\mathbf{x}^{(0)}(t)$ и определяют его устойчивость. Действительно, проекция возмущения на собственные векторы $\mathbf{Y}(T)$ через период умножается на ρ_i, и затуханию возмущения должно отвечать $| \rho_i |< 1$. Периодическому решению соответствует $\rho = 1$.

Через мультипликаторы определяются характеристические ляпуновские показатели:

$$\lambda_i = \ln | \rho_i | /T.$$

Если $| \rho_i |< 1$, то решение устойчиво. Устойчивому предельному циклу соответствует сигнатура спектра ляпуновских характеристических показателей “0, –, –, ..., –”.

Появление дополнительных мультипликаторов на границе единичного круга в комплексной плоскости свидетельствует о бифуркационной ситуации и требует дополнительных исследований.

2.4. Типичные бифуркации нелинейных динамических систем

2.4.1. Бифуркация смены устойчивости. Рассмотрим динамическую систему с фазовым пространством размерности единица, эволюция которой описывается уравнением

$$\dot{x} = F(x, \mu), \tag{2.13}$$

и пусть x_0 — состояние равновесия, т.е.

$$F(x_0, \mu) = 0. \tag{2.14}$$

Тогда, если $F'_x(x_0, \mu) \neq 0$ и функция $s(\mu) = F'_x(x_0, \mu)$ непрерывна по μ, так, что уравнение (2.14) разрешимо относительно x_0 в окрестности точки μ, то состояние равновесия x_0 является грубым, так как малым изменением параметра μ мы не изменим в качественном отношении фазовый портрет системы — точка равновесия не исчезает и новых точек равновесия не появляется. Линеаризованное уравнение для (2.13) записывается в виде

$$\frac{\partial}{\partial t}(x - x_0) = F'_x(x_0, \mu)(x - x_0),$$

и устойчивость точки x_0 определится знаком первой производной: при $F'_x(x_0, \mu) < 0$ равновесие устойчиво, а при $F'_x(x_0, \mu) > 0$ — неустойчиво. Если же при некотором значении μ_0 параметра μ производная $F'_x(x_0, \mu)$ обращается в нуль: $F'_x(x_0, \mu_0) = 0$, то при переходе через точку $\mu = \mu_0$ происходит смена устойчивого режима на неустойчивый, т.е. в точке $\mu = \mu_0$ имеет место бифуркация смены устойчивости.

Например, пусть динамическая система описывается эволюционным уравнением

$$\dot{x} = \mu \sin(x).$$

Тогда при $\mu > 0$ точка $x = 0$ задает грубое устойчивое состояние равновесия, при $\mu < 0$ точка $x = 0$ — грубое неустойчивое состояние равновесия, а при $\mu = 0$ происходит бифуркация смены устойчивости.

Заметим, что значение производной $F'_x(x_0, \mu)$ является характеристическим ляпуновским показателем динамической системы.

2.4.2. Бифуркация "седло–узел". Складка. Пусть в эволюционном уравнении (2.13) правая часть задана равенством

$$F = -\mu_1 + \mu_2 x^2,$$

и для определенности $\mu_2 > 0$. Тогда при $\mu_1 > 0$ в системе существуют два положения равновесия — устойчивое и неустойчивое; при $\mu_1 = 0$ они сливаются в одно, а при $\mu_1 > 0$ — исчезают. В комбинированном пространстве фазовой координаты и параметров $\{\mu_1, \mu_2\}$ бифуркационная диаграмма выглядит как складка, рис. 2.11; здесь координата точки поверхности, отложенная по вертикальной оси, задает положение точки равновесия системы. Бифуркация $\mu_1 = 0$ называется бифуркацией срыва равновесия или седло–узловой бифуркацией (в точке $\mu_1 = 0$ устойчивый и неустойчивый узлы сливаются в одну точку равновесия системы и исчезают при дальнейшем изменении параметра μ_1. Ситуация, когда по одному направлению возмущения система является устойчивой, а по другому — неустойчивой, называется седловой). Она также называется бифуркацией коразмерности единица, так как выделяется единственным условием: в точке бифуркации $F'_x(x_0, \mu_0) = 0$.

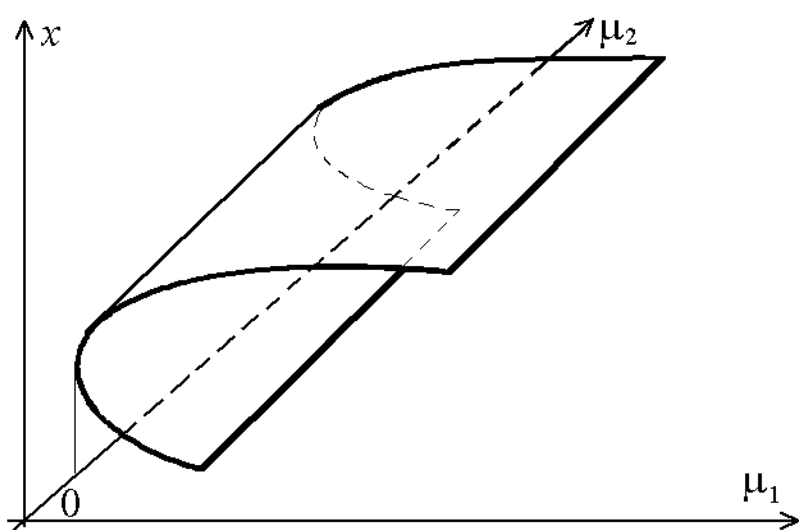


Рис. 2.11. Бифуркационная диаграмма в виде складки

2.4.3. Сборка. Рассмотрим уравнение $\dot{x} = c + \mu_1 x + \mu_2 x^3$, задающее эволюцию системы, где по-прежнему, как и в предыдущем разделе, $\mu_2 > 0$. В зависимости от значения параметра μ_1 в

системе может существовать либо три, либо одно грубое состояние равновесия.

На рис. 2.12 такая ситуация изображается поверхностью типа сборки. Изменение параметров μ_1 и μ_2 вдоль кривых (μ_1, μ_2), идущих под ненулевым углом к границе заштрихованной области — бифуркационным линиям l_0 (как говорят, трансверсально к l_0), характеризуется гистерезисом: при движении справа налево система скачком меняет свое устойчивое равновесие при значениях параметров, соответствующих точке A_1 на плоскости параметров $\{\mu_1, \mu_2\}$, а при возвращении по той же линии скачкообразное изменение устойчивого равновесия происходит в точке A_2.

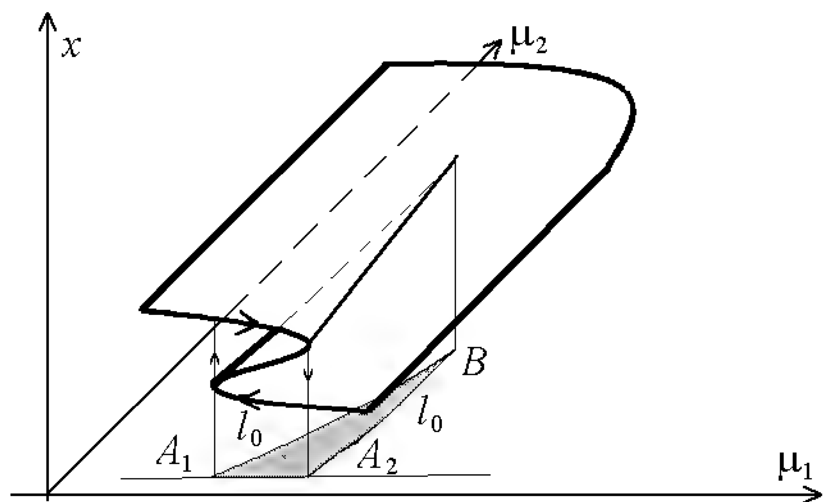


Рис. 2.12. Бифуркационная диаграмма в виде сборки

В точке B выполняются два равенства $F'_x = 0$, $F''_{xx} = 0$, поэтому говорят, что при значениях параметров, соответствующих точке B, происходит бифуркация коразмерности два.

Складка и сборка являются элементарными особенностями поверхностей, из которых может быть скомбинирована любая особенность поверхностей в трехмерном пространстве (x, μ_1, μ_2).

2.4.4. Бифуркация рождения предельного цикла. Пусть размерность фазового пространства равна двум, и эволюционное уравнение имеет вид

$$\dot{x}_1 = F_1(x_1, x_2, \mu),$$
$$\dot{x}_2 = F_2(x_1, x_2, \mu).$$

Устойчивость траектории определяется собственными числами матрицы линеаризации $A_{ij} = \partial F_i.\partial z_j$, $i, j = 1, 2$, которые могут принимать и комплексные значения. Их реальные части дают ляпуновские характеристические показатели динамической системы. Запишем вектор $\mathbf{x} = (x_1, x_2)$ в комплексной форме $z = x_1 + ix_2 = |z|e^{\arg z}$, и зададим эволюционное уравнение в комплексном виде, например, следующим образом:

$$\dot{z} = (\mu + i\omega)z + Vz \mid z \mid^2 . \tag{2.15}$$

Найдем особые точки системы, приравняв правую часть (2.15) нулю. Эта операция дает единственную точку покоя: $z_0 = 0$. Это фокус, который устойчив при $\mu < 0$ и неустойчив при $\mu > 0$.

Введем полярные координаты:

$$I = |z|^2, \quad \vartheta = \arg z. \tag{2.16}$$

Рис. 2.13. Фазовые траектории, иллюстрирующие рождение устойчивого предельного цикла

Рис. 2.14. Фазовые траектории, иллюстрирующие исчезновение неустойчивого предельного цикла

Умножим (2.15) на z^* и сложим с комплексно сопряженным уравнением. Получим

$$\dot{I} = 2I(\mu + VI).$$

Кроме положения равновесия $I_0 = 0$ (соответствующего равновесию $z = z_0 = 0$), здесь есть еще одно равновесное значение

$$I_1 = -\mu/V, \qquad (2.17)$$

которое, согласно (2.16), должно иметь знак $I_1 > 0$. Поэтому оно существует только тогда, когда параметры μ и V имеют разные знаки.

Рис. 2.15. Бифуракция удвоения периода цикла

При $I = \text{const}$ вторая переменная ϑ может изменяться с течением времени, т.е. I_1 в (2.17) определяет предельный цикл. В окрестности $I = I_1$

$$\dot{I} = 2IV(I - I_1),$$

и при $V < 0$ предельный цикл существует и устойчив при $\mu > 0$. Если же при этом $\mu < 0$, то предельный цикл отсутствует; так происходит рождение устойчивого предельного цикла из устойчивого фокуса при переходе μ через нуль с изменением от отрицательных к положительным значениям. При $V > 0$ ситуация обратная.

При переходе μ от отрицательных значений к положительным через нуль происходит исчезновение неустойчивого предельного цикла. Это уже встречавшаяся ранее бифуркация Пуанкаре–Андронова–Хопфа. Рис. 2.13, *а*, *б* изображают фазовые траектории накануне, а рис. 2.13,*в* — после рождения устойчивого предельного цикла. На рис. 2.14,*а* изображены траектории до исчезновения предельного цикла, а на рис. 2.14, *б*, *в* — после него.

2.4.5. Бифуркации удвоения периода и расщепления цикла. Фазовые траектории циклического движения удобно исследовать с помощью отображения Пуанкаре. Рассмотрим плоскость, трансверсально пересекающую фазовые траектории. Точку x_0 на этой плоскости фазовый поток переводит в точку $x_1 = F(x_0)$. Отображение F является отображением Пуанкаре. Запишем его в виде

$$x_{n+1} = F(x_n, \mu).$$

Пусть x_0 — неподвижная точка отображения Пуанкаре; это означает, что она лежит на циклической траектории. Устойчивость этой траектории определяется мультипликаторами цикла: вектор $\mathbf{e}_3$ — собственный вектор оператора монодромии, направленный по нормали к плоскости Пуанкаре, должен обладать единичным мультипликатором, эта ситуация показана на рис. 2.15.

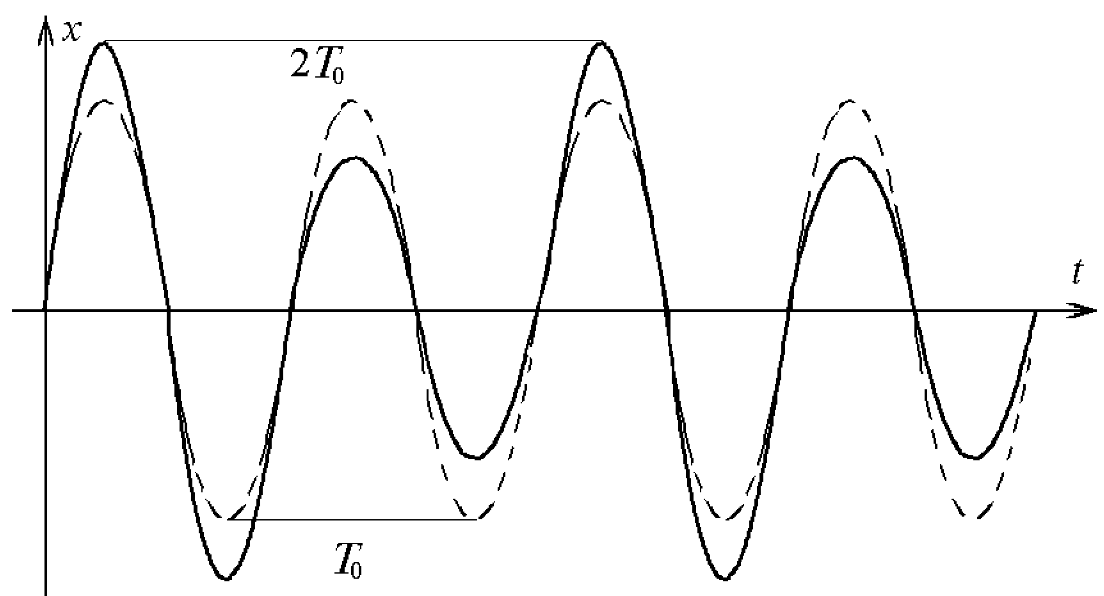


Рис. 2.16. Биения, возникающие при удвоении периода цикла

Если $| \rho_i(\mu) | < 1$, $i = 1, 2$, то любая фазовая кривая, проходящая в малой окрестности точки x_0, с течением времени будет стремиться к x_0, а если при каком-либо значении μ мультипликатор ρ_2 становится равным минус единице, то происходит удвоение периода цикла.

Вблизи точки x_0, в которой при этом остается неустойчивый (седловой) цикл, происходит рождение цикла с удвоенным периодом. Это соответствует биениям, изображенным на рис. 2.16.

Если при каком-либо значении μ еще один мультипликатор цикла

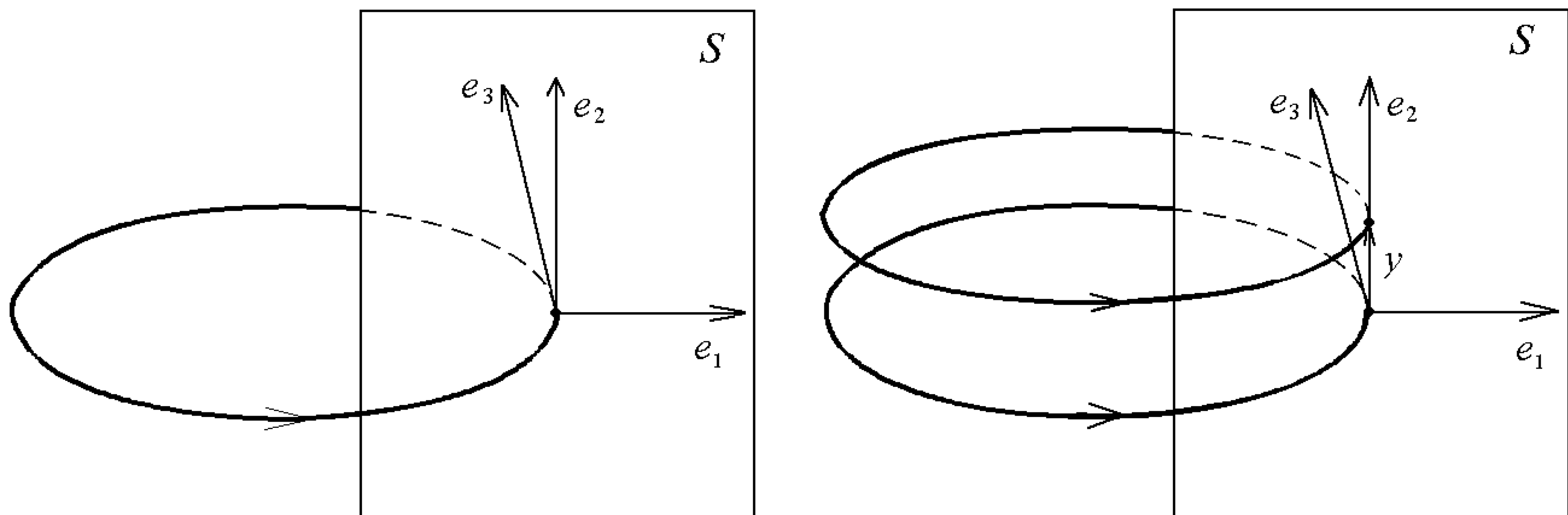


Рис. 2.17. Расщепление предельного цикла

ρ_2, помимо ρ_1, становится равным единице, то происходит расщепление цикла на два цикла, [1]. Эта ситуация, иллюстрируемая рис. 2.17, является бифуркационной, так как значение $\rho_2 = 1$ делит две области: $\rho_2 < 1$, где предельный цикл устойчив, и $\rho_2 > 1$, где предельный цикл неустойчив.

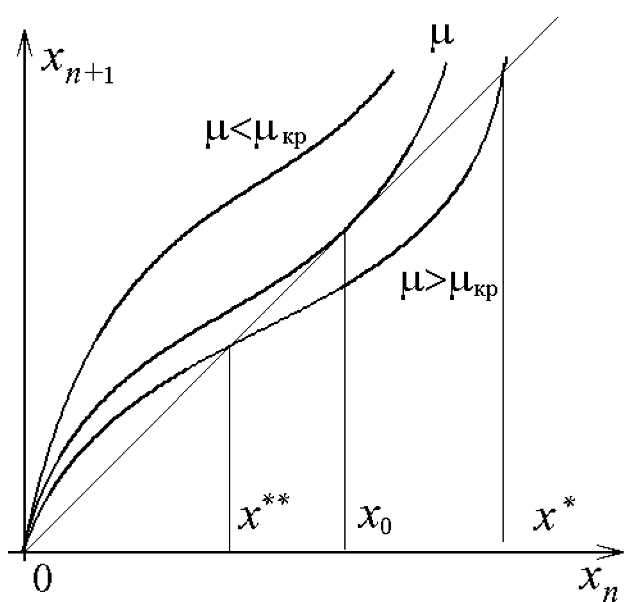


Рис. 2.18. Отображение Пуанкаре при расщеплении цикла

Расщепление предельного цикла на два цикла удобно также проиллюстрировать на языке отображения Пуанкаре. Предельному циклу соответствует неподвижная точка отображения Пуанкаре, когда $x_{n+1} = x_n$. На рис. 2.18 приведен график $x_{n+1} = F(x_n)$ для скалярной координаты x. Неподвижной точкой этого отображения, соответствующей циклу динамической системы, является точка пересечения графика функции $F(x)$ с диагональю первого координатного угла. При переходе параметра μ через критическое значение $\mu = \mu_{\text{кр}}$ происходит исчезновение или рождение предельного цикла: при $\mu = \mu_{\text{кр}}$ существует один предельный цикл $x = x_0$, который исчезает при $\mu < \mu_{\text{кр}}$ и расщепляется на два — $x = x^*$ и $x = x^{**}$ при $\mu > \mu_{\text{кр}}$.

2.5. Введение в элементарную теорию катастроф

2.5.1. Вводные замечания и примеры. Так же, как и теория бифуркаций, рассмотренная нами в предыдущем разделе, элементарная теория катастроф занимается исследованием поведения динами-

ческих систем в окрестности точки равновесия в зависимости от значений управляющих параметров. Однако, в отличие от подходов, рассмотренных выше, где исследовались лишь линейные члены в разложении правой части эволюционных уравнений вокруг точки равновесия, определяющие устойчивость системы, теория катастроф интересуется и тем, каким образом состояния равновесия изменяются при изменении управляющих параметров, при этом особое внимание уделяется ситуациям, в которых сингулярные числа матрицы линеаризации обращаются в нуль, и приходится учитывать члены более высокого порядка, чем линейные — квадратичные, кубические и т.п. При этом указывается, что в ряде случаев всевозможные возмущения с помощью гладкой замены переменных можно свести к некоторому набору "канонических", и тем самым удается разделить их на эквивалентные классы, такие, что представителям одного и того же класса соответствует качественно неразличимые поведения динамических систем. Благодаря этим результатам, вместо изучения всего разнообразия возмущений точек равновесия имеет смысл исследовать лишь небольшой набор их канонических представителей.

Рассмотрим подходы элементарной теории катастроф, пользуясь для иллюстрации передемпфированной системы материальных точек, находящихся в вязкой среде, скорость движений которых пропорциональна действующей силе. Пример такой системы приведен на предыдущей лекции. Эволюционное уравнение в такой ситуации имеет вид

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x};\mathbf{c}) = \frac{\partial U(\mathbf{x};\mathbf{c})}{\partial \mathbf{x}}, \tag{2.18}$$

где вектор $\mathbf{x}$ описывает состояние динамической системы (координаты материальных точек), $\mathbf{c}$ — вектор управляющих параметров, t — время, U — потенциал силового поля $\mathbf{F}$. В элементарной теории катастроф правая часть (2.18) с помощью гладкой замены переменных приводится к одному из канонических видов, и тем самым происходит классификация динамических систем с точки зрения их возможного качественного поведения в точках равновесия (особых точках) в зависимости от числа координат вектора $\mathbf{x}$ и вектора управляющих параметров $\mathbf{c}$.

Поясним на нескольких примерах, как может быть осуществлена такая классификация.

Пример 1. Пусть $\mathbf{x}$ — точка, отличная от состояния равновесия, т.е. в этой точке $\nabla U \neq 0$. Тогда с помощью гладкой замены переменных в окрестности этой точки движение системы можно описать как движение вдоль направления градиента потенциала U. Система качественно будет эквивалентна одномерной системе, описываемой уравнением

$$\dot{x}_1 = 1,$$

т.е. потенциал силового поля в этой случае может быть приведен к виду $U = x_1 + \text{const}$.

Пример 2. Пусть $\mathbf{x}$ — точка равновесия; в ней выполнено соотношение $\nabla U = 0$, и движение определяется значениями вторых производных $A_{ij} = \partial^2 U/\partial x_i \partial x_j$ потенциала U в рассматриваемой точке. В общем случае собственные значения симметричной матрицы A ненулевые, и в окрестности рассматриваемой точки выражение для потенциала можно записать в виде

$$U = \sum_{i=1}^{N} \lambda_i x_i^2 + O\left(\|x\|^3\right), \tag{2.19}$$

однако гладкой заменой переменных $y_i = y_i(x_1, \ldots, x_N)$, $i = 1, \ldots, N$, можно избавиться от степеней выше второй в (2.19) так, что движение системы в окрестности рассматриваемой точки будет эквивалентно движению в силовом поле квадратичного потенциала вблизи его точки экстремума; потенциал силового поля при этом будет иметь вид

$$U = \sum_{i=1}^{N} \lambda_i x_i^2. \tag{2.20}$$

Точки равновесия, в которых возможно представление потенциала в виде (2.20), называются морсовскими точками.

2.5.2. Неморсовские особые точки. Росток и возмущение катастрофы. Морсовская точка равновесия наиболее интересна с той точки зрения, что в этой точке экстремума функции U можно так распорядиться k управляющими параметрами, чтобы $l \leq k$ собственных значений матрицы A сделать равными нулю (при этом точка перестанет быть морсовской), и тогда эквивалентный потенциал не будет квадратичным. Рене Том исследовал этот вопрос, и предложил классификацию неморсовских особых точек в зависимости от числа нулевых собственных значений, числа управляющих параметров и др. Точки пространства параметров (значения управляющих параметров), в которых детерминант матрицы A обращается в нуль, называют точками катастроф.

Общий вид потенциала в окрестности неморсовских особых точек, или точек катастроф, согласно Р.Тому, можно записать в виде

$$U = G(x_1, \ldots, x_l) + \sum_{i=l+1}^{N} \lambda_i x_i^2, \tag{2.21}$$

где функция $G(x_1, \ldots, x_l)$ называется ростком катастрофы. С точки зрения и теории, и приложений интересно поведение потенциала

вблизи точки катастрофы не только в зависимости от координат $\mathbf{x} = (x_1, \ldots, x_N)$, как это записано в формуле (2.21), но и в зависимости от параметров $\mathbf{c} = (\mu_1, \ldots, \mu_k)$. Поэтому в $N + k$-мерном пространстве $\mathcal{R}_N \otimes \mathcal{R}_k$, являющимся прямым произведением пространства координат на пространство параметров, в окрестности особой точки катастрофы $(\mathbf{x}_0, \mathbf{c}_0)$, потенциал U может быть приведен к виду

$$U(\mathbf{x}, \mathbf{c}) = \mathrm{Cat}\,(x_1, \ldots, x_l, \mu_1, \ldots, \mu_k) + \sum_{i=l+1}^{N} \lambda_i(\mathbf{c}) x_i^2, \tag{2.22}$$

где

$$\mathrm{Cat}\,(x_1, \ldots, x_l, \mu_1, \ldots, \mu_k) = G(x_1, \ldots, x_l) + \mathrm{Pert}\,(x_1, \ldots, x_l, \mu_1, \ldots, \mu_k),$$

и $\mathrm{Cat}\,(\cdot, \ldots, \cdot)$ называется функцией катастрофы, а $\mathrm{Pert}\,(\cdot, \ldots, \cdot)$ — функцией возмущения катастрофы.

2.5.3. Классификация катастроф. Одним из основных достижений теории катастроф является доказательство того факта, что как функция ростка катастрофы $G(\cdot, \ldots, \cdot)$, так и функция ее возмущения $\mathrm{Pert}\,(\cdot, \ldots, \cdot)$ имеют канонический вид, к которому с помощью гладкой замены переменных может быть приведен потенциал любой динамической системы вида (2.18) в окрестности точки катастрофы. Классификация катастроф и канонический вид ростка и возмущения катастрофы приведен в табл 2.1.

Таблица 2.1

Тип катастрофы	k	l	Росток	Возмущение
A_2	1	1	x^3	$a_1 x$
$A_{\pm 3}$	2	1	$\pm x^4$	$a_1 x + a_2 x^2$
A_4	3	1	x^5	$a_1 x + a_2 x^2 + a_3 x^3$
$A_{\pm 5}$	4	1	$\pm x^6$	$a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4$
$D_{\mp 4}$	3	2	$x^2 y \mp y^3$	$a_1 x + a_2 y + a_3 y^2$
D_5	4	2	$x^2 y + y^4$	$a_1 x + a_2 y + a_3 x^3 + a_4 y^2$
$D_{\pm 6}$	5	2	$x^2 y \pm y^5$	$a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 y^3$

Поясним, как получена табл 2.1. Рассмотрим замену переменных

$$\mathbf{x}' = \mathbf{x}'(\mathbf{x})$$

с отличным от нуля якобианом

$$\det \left| \frac{\partial x_i'}{\partial x_j} \right| \neq 0.$$

Разложим функции $\mathbf{x}'(\cdot)$ в ряд Тейлора в точке $\mathbf{x}_0$ и запишем

$$\mathbf{x}'(\mathbf{x}) = A_1 + \widehat{A}_2\mathbf{x} + A_3(\mathbf{x}),$$

где $A_1 \in \mathcal{R}_n$ — константа, $\widehat{A}_2 \in (\mathcal{R}_n \to \mathcal{R}_n)$ — линейное преобразование, а A_3 — так называемое осесохраняющее нелинейное преобразование вида

$$(A_3(\mathbf{x}))_i = \sum_i \left(\delta_{ij}x_j + \sum_k \left(A_{i,jk}x_jx_k + \sum_l (A_{i,jkl}x_jx_kx_l + \ldots)\right)\right),$$

и покажем, как с помощью этих преобразований можно преобразовать эволюционные уравнения к простейшему каноническому виду.

2.5.4. Канонический вид эволюционного уравнения в неособой точке. Рассмотрим точку $\mathbf{x}$, в которой $\nabla U \neq 0$. Так как, согласно эволюционному уравнению (2.18), в результате действующей силы возникает движение, то любые две потенциальные функции, описывающие в некоторой точке отличные от нуля силы, должны быть качественно одинаковыми в этой точке. Рассмотрим преобразование

$$\begin{aligned} x_1' &= U(x_1, x_2, \ldots, x_N), \\ \mathbf{x}_2' &\equiv (x_2', \ldots, x_N') = A\mathbf{x}. \end{aligned} \tag{2.23}$$

Якобиан этого преобразования равен

$$\det\left|\frac{\partial x_i'}{\partial x_j}\right| = \det\begin{vmatrix} \dfrac{\partial U}{\partial x_1} & \dfrac{\partial U}{\partial x_2} & \ldots & \dfrac{\partial U}{\partial x_N} \\ a_{21} & a_{22} & \ldots & a_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ a_{N1} & a_{N2} & \ldots & a_{NN} \end{vmatrix} \tag{2.24}$$

и может быть сделан отличным от нуля в точке $\mathbf{x}_0 = 0$ выбором матрицы A, если

$$\frac{\partial U}{\partial x_j} \neq 0, \quad j = 1, \ldots, N.$$

Тогда $U'(\mathbf{x}') = x_1' = U(x_1, \ldots, x_n) = U(\mathbf{x})$, и поведение штрихованной и нештрихованной динамических систем эквивалентны.

2.5.5. Канонический вид эволюционного уравнения в морсовской точке. Лемма Морса. Пусть теперь $\nabla U = 0$, тогда первая строка (2.24) обращается в нуль и преобразование (2.23) — необратимо (не является гладким). Поскольку рассматриваемая точка морсовская, то матрица вторых производных потенциала $\left(\partial^2 U/\partial x_i\partial x_j\right)$ невырождена, т.е. не имеет равных нулю собственных значений. Покажем,

что в этом случае выбором неособого преобразования координат $x_1, \ldots, x_N$ можно привести потенциал к квадратичному виду. Пусть для примера $N = 2$ и в точке, где $\nabla U = 0$ (не ограничивая общности можно считать, что это — начало координат) можно записать по формуле Тейлора

$$U(x, y) = ax^2 + bxy + cy^2 + \text{члены третьей степени} + \ldots$$

Линейным преобразованием приведем квадратичные члены к диагональному виду:

$$U(\widetilde{x}, \widetilde{y}) = \lambda_1 \widetilde{x}^2 + \lambda_2 \widetilde{y}^2 + (d\widetilde{x}^3 + e\widetilde{x}^2\widetilde{y} + f\widetilde{x}\widetilde{y}^2 + g\widetilde{y}^3) + \ldots \tag{2.25}$$

Далее определим нелинейное преобразование

$$\begin{aligned} x' &= \widetilde{x} + \Delta_1 = \widetilde{x} + (A_{20}\widetilde{x}^2 + A_{11}\widetilde{x}\widetilde{y} + A_{02}\widetilde{y}^2) + \ldots\ , \\ y' &= \widetilde{y} + \Delta_2 = \widetilde{y} + (B_{20}\widetilde{x}^2 + B_{11}\widetilde{x}\widetilde{y} + B_{02}\widetilde{y}^2) + \ldots \end{aligned} \tag{2.26}$$

и выберем A_{pq}, B_{pq}, $p + q \geq 2$, так, чтобы привести потенциал U к канонической морсовской форме:

$$U'(x', y') = \lambda_1 {x'}^2 + \lambda_2 {y'}^2. \tag{2.27}$$

Подставим (2.26) в (2.27) и с учетом (2.25) получим

$$\lambda_1 \left(\widetilde{x} + \sum_{p+q \geq 2} A_{pq} \widetilde{x}^p \widetilde{y}^q \right)^2 + \lambda_2 \left(\widetilde{y} + \sum_{p+q \geq 2} B_{pq} \widetilde{x}^p \widetilde{y}^q \right)^2 = \\ = \lambda_1 \widetilde{x}^2 + \lambda_2 \widetilde{y}^2 + \left(d\widetilde{x}^3 + e\widetilde{x}^2\widetilde{y} + f\widetilde{x}\widetilde{y}^2 + g\widetilde{y}^3 \right) + \ldots$$

Приравнивая коэффициенты при одинаковых степенях $\widetilde{x}$ и $\widetilde{y}$, найдем систему уравнений для определения коэффициентов нелинейного преобразования (2.26):

$$\begin{aligned} x^3 :&\quad \lambda_1 2A_{20} = d, \\ x^2 y :&\quad \lambda_1 2A_{11} + \lambda_2 2B_{20} = e, \\ xy^2 :&\quad \lambda_1 2A_{02} + \lambda_2 2B_{11} = f, \\ y^3 :&\quad \lambda_2 2B_{02} = g. \end{aligned} \tag{2.28}$$

Уравнений (2.25)– (2.27) достаточно для определения коэффициентов A_{pq}, B_{pq}, $p + q \geq 2$. Действительно, на шесть коэффициентов A_{pq}, B_{pq}, $p + q = 2$, наложены лишь ограничения (2.28), состоящие из четырех уравнений с четырьмя заданными числами d, e, f, g.

Аналогично могут быть найдены коэффициенты при более высоких степенях $\widetilde{x}$, $\widetilde{y}$.

Итак, в морсовской точке поведение любой системы эквивалентно динамике системы, описываемой квадратичным потенциалом. Система же с квадратичным потенциалом либо устойчива к любым возмущениям из $\mathcal{R}_N$, если все $\lambda_i < 0$, либо неустойчиво по всем направлениям в $\mathcal{R}_N$, если все $\lambda_i < 0$, $i = 1, \ldots, N$, либо, если $\lambda_1 < 0, \ldots, \lambda_p < 0, \lambda_{p+1} > 0, \ldots, \lambda_N > 0$, $1 < p < N$, то в $\mathcal{R}_N$ есть как устойчивые, так и неустойчивые направления возмущений (особенность типа седла).

2.5.6. Канонический вид эволюционного уравнения в особой точке катастроф. Возмущения. Коэффициенты квадратичной формы, задающей потенциал в окрестности морсовской особой точки, зависят от управляющего параметра $\mathbf{c} \in \mathcal{R}_k$, и можно так распорядиться значениями координат вектора $\mathbf{c}$, чтобы обратить в нуль часть этих коэффициентов. В общем невырожденном случае одной координаты вектора $\mathbf{c}$ достаточно, чтобы обнулить одно собственное значение матрицы $\partial^2 U/\partial x_i \partial x_j$; канонический вид потенциала тогда должен содержать более высокие степени соответствующих координат (кубические и т.д.). Рассмотрим этот вопрос подробнее и найдем вид зависимости потенциала от параметров $\mathbf{c}$ и координат вектора $\mathbf{x} \in \mathcal{R}_N$ в окрестности неморсовской особой точки.

В морсовских точках выполнено равенство

$$\det \left| \frac{\partial^2 U}{\partial x_i \partial x_j} \right| = 0,$$

и значит, одно или несколько собственных значений λ_i, $i = 1, \ldots, l$, $1 \le l \le k$, матрицы вторых производных потенциала равны нулю. Пусть, для примера, в рассматриваемом двумерном случае $\lambda_1 = 0$. Тогда из (2.28) можно найти коэффициенты B_{02}, B_{11}, B_{20}, тем самым избавиться в канонической формуле для потенциала U' от кубических степеней $x^2 y$, xy^2, y^3, но исключить из разложения x^3 не удается. Написав разложение для четвертых степеней, нетрудно заметить, что не удается исключить из разложения потенциала слагаемые с x^4, и т.д. Таким образом, применяя описанную выше процедуру, придем к представлению

$$U'(x', y') = px'^3 + qx'^4 + \ldots + \lambda_2 y'^2 = f_{MN}(x') + \lambda_2 y'^2.$$

Здесь, однако, пока мы не использовали возможности нелинейного преобразования координаты $x' \to x''$ по формуле

$$x'' = A_1 x' + A_2 x'^2 + \ldots$$

Повторяя ход рассуждения предыдущего раздела, можно показать, что при такой замене придем к канонической форме катастрофы A_2 (см. табл 2.1):

$$U''(x'', y'') = {x''}^3 + \lambda_2 {y''}^2,$$

здесь $y'' = y'$.

Если в динамической системе имеется более, чем один управляющий параметр, то оставшиеся (после обнуления λ_1) $k-1$ параметр можно использовать для того чтобы избавиться в разложении (2.25) от более высоких степеней x, в результате получится катастрофа вида A_{k+1}. Канонический вид возмущения катастроф (последний столбец табл. 2.1) получается аналогично. Здесь изменение параметров, т.е. движение вдоль вектора $(0, \mathbf{c})$ в пространстве $\mathcal{R}_N \otimes \mathcal{R}_k$, позволяет добавить к ростку катастрофы столько слагаемых полиномиального вида, сколько параметров остаются свободными.

Итак, основную идею элементарной теории катастроф можно сформулировать следующим образом: при разложении потенциала по формуле Тейлора в окрестности исследуемой особой точки управляющие параметры системы используются, чтобы обратить в нуль коэффициенты при младших степенях полинома Тейлора (меньших, чем в каноническом виде функции катастрофы), а нелинейное невырожденное преобразование координат позволяет избавиться от старших степеней полинома (превосходящих степень полинома канонического представления).

2.5.7. Сепаратрисы на множестве параметров. Рассмотрим семейство "типичных" потенциальных функций $U(\mathbf{x}; \mathbf{c})$, $\mathbf{x} \in \mathcal{R}_n$, $\mathbf{c} = (\mu_1, \ldots, \mu_k) \in \mathcal{R}_k$. Если точка $\mathbf{c}^0 \in \mathcal{R}_k$ — "случайная", или "типичная", то функция $U(\mathbf{x}; \mathbf{c})$ имеет лишь морсовские критические точки, неморсовские точки характерны лишь для параметров из множества меры нуль пространства параметров $\mathcal{R}_k$. Однако в основном неморсовские критические точки ответственны за качественную природу семейства потенциальных функций. Множество точек $\mathbf{c} \in \mathcal{R}_k$, в которых $U(\mathbf{x}, \mathbf{c})$ являются неморсовскими (т.е. имеют хотя бы одну неморсовскую точку), называются сепаратрисами. Заметим, что на предыдущих лекциях нами рассматривались сепаратрисы на фазовых плоскостях, здесь же идет речь о сепаратрисах на множестве внешних параметров.

Найдем некоторые сепаратрисы.

Катастрофа типа A_2. В этом случае $U(\mathbf{x}, \mathbf{c})$ — однопараметрическое семейство функций. Т.к. на морсовскую часть катастрофа не влияет, рассмотрим только неморсовскую часть, которая, согласно табл. 2.1, имеет вид $F(x, a) = x^3/3 + ax$. Здесь ростком катастрофы является функция $G = \frac{1}{3}x^3$, а возмущением — функция $\mathrm{Pert} = ax$. Точка $a = 0$ является сепаратрисой, так как при $a = 0$ точка $x = 0$ является неморсовской. При $a < 0$ все функции вида $F(x, a) = \frac{1}{3}x^3 + ax$

являются функциями качественно одного и того же типа, так как все они в окрестности точки $x = 0$ имеют две критические точки $x_{1,2} = \pm\sqrt{a}$. Аналогично, те же функции при $a > 0$ качественно не различаются потому, что не имеют критических точек. Таким образом, точка $a = 0$ делит числовую прямую на две области, отличающиеся тем, что значения управляющих параметров из разных областей приводят к различному поведению динамической системы в окрестности точки $x = 0$, рис. 2.19.

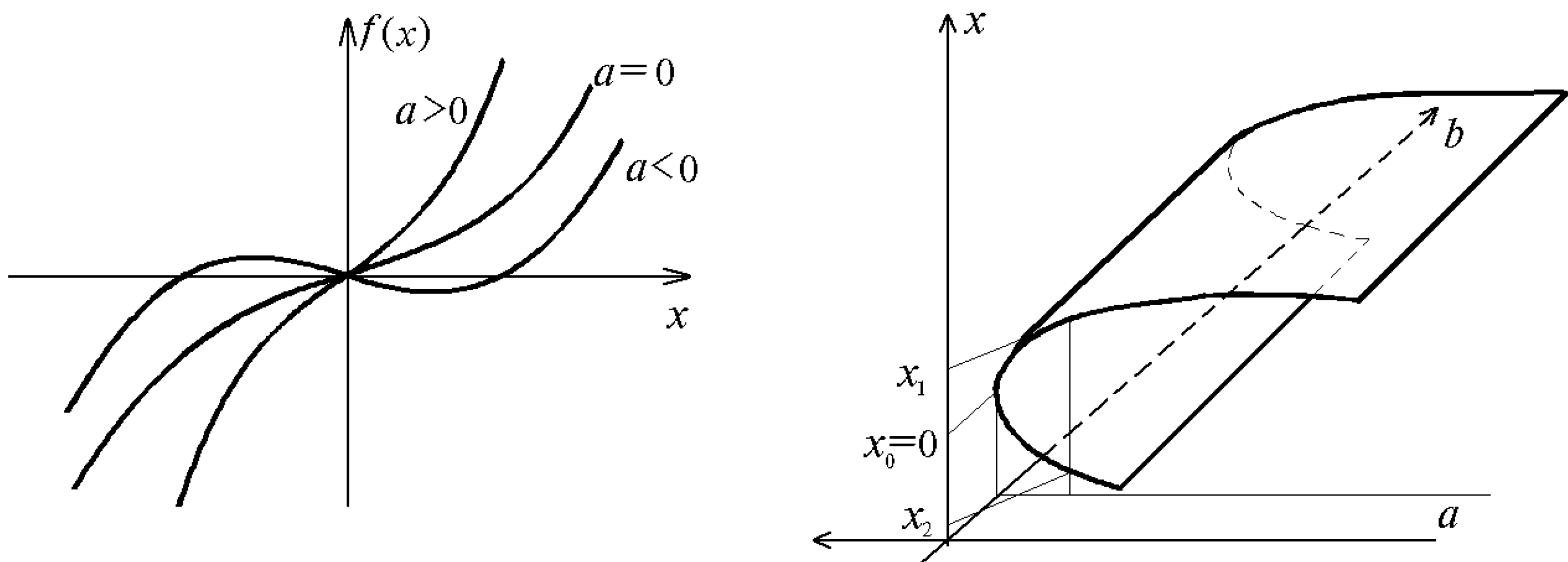


Рис. 2.19. Точка a=0 – сепаратриса на множестве параметров

Катастрофа типа A_2 рассматривалась на другом языке в предыдущем разделе, когда речь шла о седло–узловой бифуркации — в точке бифуркации две особые точки, устойчивая и неустойчивая, сливались в одну и исчезали при дальнейшем изменении параметра. Образом этой бифуркации является складка, так как цилиндрическая поверхность складки, ординаты точек которой задают положение особых точек динамической системы, по сути, описывается одном управляющим параметром; при проходе через границу складки две особые точки тоже "схлопываются" в одну (на ребре складки) и затем исчезают.

Катастрофа типа A_3. В этом случае функция катастрофы равна

$$F(x, a, b) = \frac{1}{4}x^4 + \frac{1}{2}ax^2 + bx$$

с ростком катастрофы $G = x^4/4$ и возмущением Pert $= ax^2/2 + bx$. Сепаратриса состоит из точки $(a, b) = (0, 0)$ — точки сборки на рис. 2.20; на языке предыдущего раздела это бифуркация коразмерности два, и "линии складки", описываемой уравнением $(a/3)^3 + (b/2)^2 = 0$ — это множество точек бифуркации коразмерности единица, изображаемых складкой на том же рисунке.

2.5.8. Флаги катастроф. Формальный ответ на вопрос о том, находится ли изучаемая система в пространстве параметров вблизи

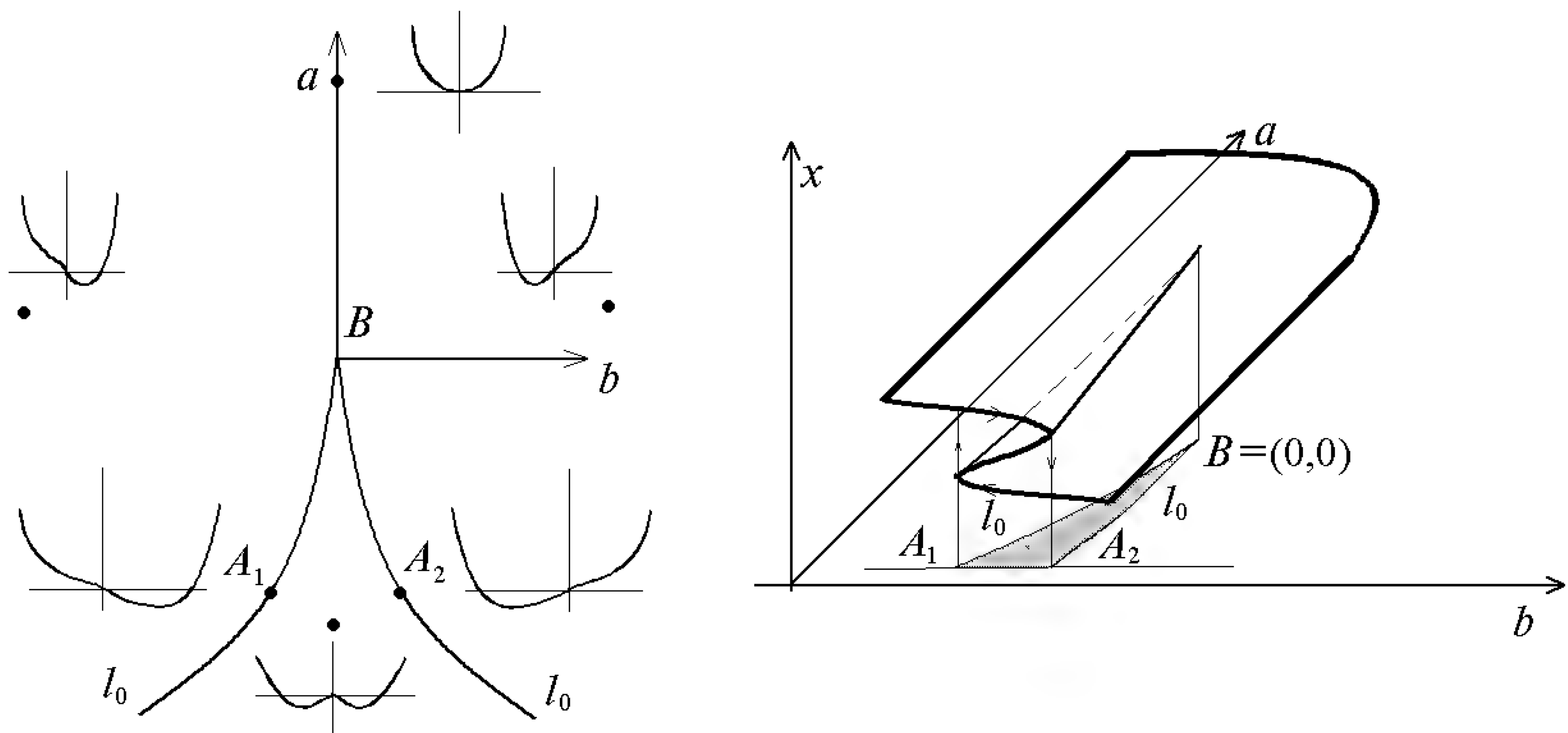


Рис.2.20. Сепаратрисы на области параметров – границы заштрихованной области. Угловая точка – катастрофа A_3, остальные точки – катастрофы A_2

некоторой точки катастрофы, состоит в исследовании зависимости потенциала от параметров и выяснении, существует ли вблизи данной точки $(\mathbf{x}, \mathbf{c})$ неморсовская точка. Однако на практике неморсовские точки по ряду причин не могут быть распознаны сразу, например, если потенциальная функция является слишком сложной или точно неизвестной. Кроме того, система может оказаться неградиентной, т.е. ее эволюционное уравнение не входит в класс уравнений, описываемых соотношением (2.18); еще хуже, когда у исследователя нет даже туманных соображений о том, какое эволюционное уравнение описывает динамику системы. Тем не менее даже в этих патологических с точки зрения математического моделирования ситуациях можно указать некоторые косвенные признаки того, что изучаемая система находится вблизи точки катастрофы. Речь идет о так называемых флагах катастроф — особенностях поведения системы, по которым можно судить о приближении катастрофы. Мы остановимся на 5 флагах, которые, как правило, встречаются вместе и "вывешиваются", когда физические параметры изменяются в некоторой окрестности, в которой имеются несколько точек минимума потенциальной функции. Если исследователь наткнулся на один из этих пяти флагов, то управляющие параметры можно поменять так, чтобы стало возможным обнаружить и другие флаги, которые обязательно должны проявить себя в подходящих условиях.

Флаги катастроф чрезвычайно полезны на практике, так как обнаружение катастрофы полезно потому, что позволяет определить упрощенную модель динамической системы, зависящую только от существенных переменных — координат и управляющих параметров. Кроме того, в ряде случаев установление факта наличия катастрофы может сделать ненужным вообще подробное математическое описание системы, если выводы определяются канонической геометрией

рассматриваемого типа катастрофы.

Опишем эти пять флагов.

1. *Модальность.* Система имеет несколько устойчивых состояний равновесия — реализуемых физических состояний. Эти состояния соответствуют наличию нескольких локальных минимумов функции потенциала U.

2. *Недостижимость.* Если у системы существует неустойчивое состояние равновесия, из которого она выводится сколь угодно малой флуктуацией, т.е. физически недостижимое состояние, то это свидетельствует о наличии в системе морсовской особой точки типа седла, а также о возможности существования двух окрестных минимумов потенциала. И наоборот, всякий раз, когда система имеет несколько локальных минимумов, то в ней обязательно должны существовать и седловые точки.

3. *Катастрофические скачки.* Если происходит резкое изменение поведения динамической системы при малом изменении параметров — катастрофический скачок, то это свидетельствует о наличии двух локальных минимумах потенциала, таких, что переход от одного из них к другому осуществляется согласно диаграмме сборки при движении по линии $A_1 - A_2$.

4. *Расходимость.* Конечные изменения параметров приводят сколь угодно большим изменениям в значении переменных, описывающих систему. Такое поведение возможно в окрестности неморсовской особой точки.

5. *Гистерезис.* Этот флаг имеет место, когда динамический процесс не является полностью обратимым. Например, при плавном изменении параметров вдоль линии $A_1 - A_2$ система скачком переходит из одного состояния в другое при значении параметров, отвечающих точке A_1, а при движении в обратном направлении — при параметрах, соответствующих точке A_2.

Флаги катастроф ярко проявляются на таких известных примерах, как фазовые переходы в системе "газ–жидкость", выгибание упругой балки под нагрузкой и др.

2.5.9. Фазовый переход как катастрофа.

Уравнение состояния реального газа (Уравнение Ван дер Ваальса). Реальный газ представляет собой сложную динамическую систему материальных тел (молекул). Если записать совокупность уравнений движения каждой молекулы, то такая модель окажется слишком сложной и громоздкой; практически ее невозможно исследовать.

Однако можно пойти по иному пути — найти некоторые параметры, характеризующие коллективное поведение системы — такими параметрами являются температура как средняя кинетическая энергия молекул и давление как средняя сила, действующая со стороны молекул на единичную площадку, плотность как среднее число молекул в

единице объема. Эти параметры с учетом конечных размеров молекул и действующих между ними сил можно записать в виде уравнения

$$\left(p + \frac{N^2 a}{U^2}\right)(U - Nb) = NkT,$$

где a и b — константы, описывающие притяжение и отталкивание молекул, $N^2 a/U^2$ — так называемое внутреннее давление, температура T является управляющим параметром. Это уравнение было предложено голландским физиком Ван дер Ваальсом в 1873 г. для описания отклонения свойств реального газа от идеального. Несмотря на то, что это уравнение является весьма приближенным и служит для количественного определения свойств реального газа лишь для областей высоких температур и низких давлений, однако качественно оно позволяет описывать поведение газа и при высоких давлениях, и конденсацию газа, и критическое состояние. Действительно, уравнение Ван-дер-Ваальса связывает давление реального газа с величиной занимаемого им объема при различных значениях температуры; график этой зависимости выглядит так, как показано на рис. 2.21, и имеет особенность типа "сборка". Благодаря этому, в окрестности точки катастрофы типа A_3, соответствующей критическому состоянию газа (точка B на плоскости (p, T)), реальный газ как физическая система проявляет все флаги катастроф: модальность как возможность существования перегретой жидкости и переохлажденного газа при одном и том же значении давления и температуры, недостижимость как неустойчивость состояния перегретой жидкости и переохлажденного пара, катастрофические скачки как мгновенное закипание перегретой жидкости или мгновенная конденсация переохлажденного газа, расходимость как те же переходы "жидкость–газ" при малых изменениях параметров — температуры или давления, гистерезис как переход от жидкости к газу через состояние перегретой жидкости и от газа к жидкости через состояние переохлажденного газа. Эти особенности катастрофы A_3 наглядно иллюстрируются диаграммой сборки.

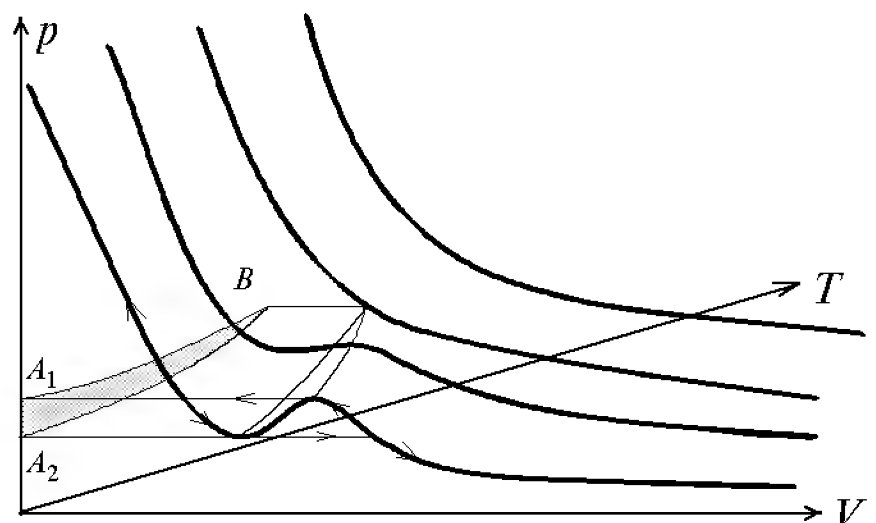


Рис. 2.21. Поверхность, описывающая возможные состояния реального газа

Балка под нагрузкой. Многие технические конструкции могут быть описаны с помощью потенциальной функции, минимальное значение которой описывает локально устойчивое состояние конструкции. С увеличением нагрузки на конструкцию потенциальная функция из-

меняется, и значительная нагрузка может привести к потере устойчивости конструкции. Это можно описать как переход от морсовской точки к неморсовской при изменении управляющего параметра — внешней нагрузки. В качестве примера рассмотрим выпучивание идеального несжимаемого стержня под действием нагрузки. Если нагрузка невелика, в системе имеется одна устойчивая точка равновесия — стержень остается прямым, рис. 2.22.

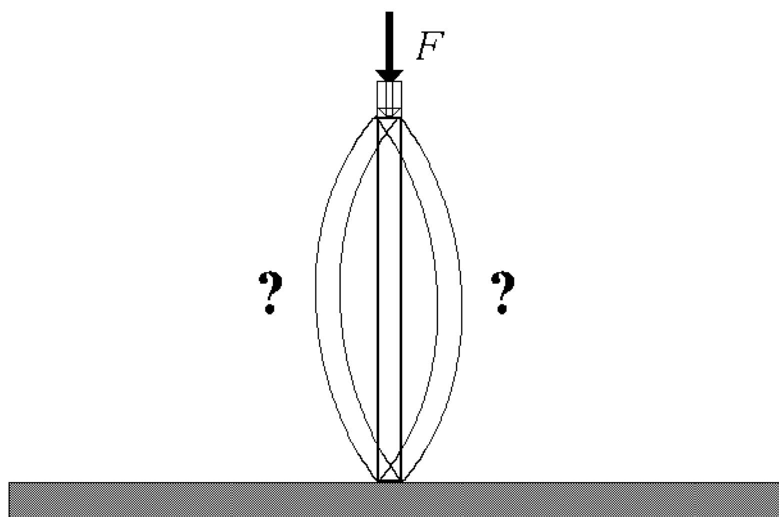


Рис 2.22. При достижении определенного усилия балка резко прогибается в ту или иную сторону

При увеличении нагрузки неизогнутое состояние стержня становится неустойчивым, и возникают два новых устойчивых состояния, изображенные на рис. 2.22. В этой ситуации также проявляются все перечисленные флаги катастроф: модальность как возможность изгиба направо или налево, недостижимость неизогнутого состояния при нагрузке, превышающей критическую, катастрофические скачки и расходимость как резкие изгибы балки при увеличении нагрузки; гистерезис проявляется в том, что при увеличении и снятии нагрузки резкие изменения формы балки от неизогнутой к изогнутой и обратно происходят при различных значениях нагрузки. Геометрическим образом описанного здесь процесса является сборка.

2.5.10. Точки катастроф и изменение климата. Согласно геологическим данным, поверхность Земли хранит следы многих оледенений, когда огромные пространства суши покрывались толстым слоем льда, причем их длительность значительно превышала продолжительность межледниковых периодов: в течение последнего миллиона лет оледенения наступали примерно каждые 100 тысяч лет, и длительность межледниковых периодов составляла 10–12 тысяч лет. Резкие изменения климата Земли влекут глобальные изменения и в растительном и животном мире, и в социальных явлениях. Поэтому важно знать, произойдут ли и в какое время колебания или изменения климата в будущем. Дело в том, что межледниковый период, в котором мы живем, начался около 10 тысяч лет назад, и есть все основания полагать, что переход к ледниковому периоду может последовать очень быстро, в течение каких-либо 100 лет, а может, и еще быстрее.

Относительное спокойствие людей, считающих стабильность климата Земли естественным ее свойством, основано на представлении о линейных системах, в поведении которых не бывает резких скачков. Однако система “поверхность Земли–атмосфера” является сложной

незамкнутой нелинейной системой, в которой возможны точки катастроф. Рассмотрим некоторые представления о ритмах ледниковых периодов и выясним сначала, каким может быть механизм возникновения ледниковых периодов.

Солнце является неравновесной системой, которая, по-видимому, находится в состоянии устойчивого равновесия. Однако эта система совершает колебания, прослеживающиеся в наблюдениях за солнечной активностью — это известные 27-суточные, 22-летние и другие циклы. Возможно, что один из периодов колебаний соизмерим со статысячелетним периодом смены ледниковых эпох. Тогда изменения интенсивности солнечного излучения могли бы влиять на ритм ледниковых периодов, если амплитуда колебаний солнечного излучения достаточно велика. Другим фактором, влияющим на климат, может оказаться космическая пыль, через места скопления которой Солнце может периодически проходить при своем движении в Галактике: наличие пыли между Солнцем и Землей ослабляет потоки солнечного излучения, достигающие поверхности Земли. При крупных извержениях вулканов также уменьшается поток тепла, приходящий на поверхность Земли — известны случаи, когда после вулканических извержений настолько увеличивалась содержание аэрозольных частиц в стратосфере, отражающих солнечный свет, что на всей Земле не наступало лето. Смена магнитного поля Земли также влияет на химический состав верхней атмосферы, что может привести к глобальным климатическим изменениям. Наконец, отметим и изменение геометрии орбиты Земли. Таким образом, можно констатировать, что имеется достаточное число факторов, которые могут влиять на климат Земли и приводить к его существенному изменению.

Одна из широко распространенных гипотез влияния параметров орбиты на климат Земли принадлежит М.Миланковичу. Он показал, что изменения эксцентриситета, наклонения эклиптики и прецессии приводят к изменению количества солнечной радиации, поступающей в верхние слои атмосферы, и могут приводить к изменению климата. Если бы Земля двигалась в поле центральных сил, то ее орбита имела бы форму эллипса, плоскость эллипса называется эклиптикой, а ее форма (вытянутость) характеризуется эксцентриситетом. Однако движение Земли вокруг Солнца возмущается другими планетами, в результате чего, в основном, изменяется эксцентриситет. За последний миллион лет его значение изменилось от 0.00 до 0.06, причем это изменение не описывается какой-либо одной частотой, а имеется спектр частот с максимумами интенсивности в районе 100 тысяч лет. Вариации количества энергии, поступающей в атмосферу, при таком изменении эксцентриситета составляет примерно 0.1%, что достаточно для изменения средней температуры на несколько градусов Кельвина. Наклон эклиптики в основном определяет межсезонные

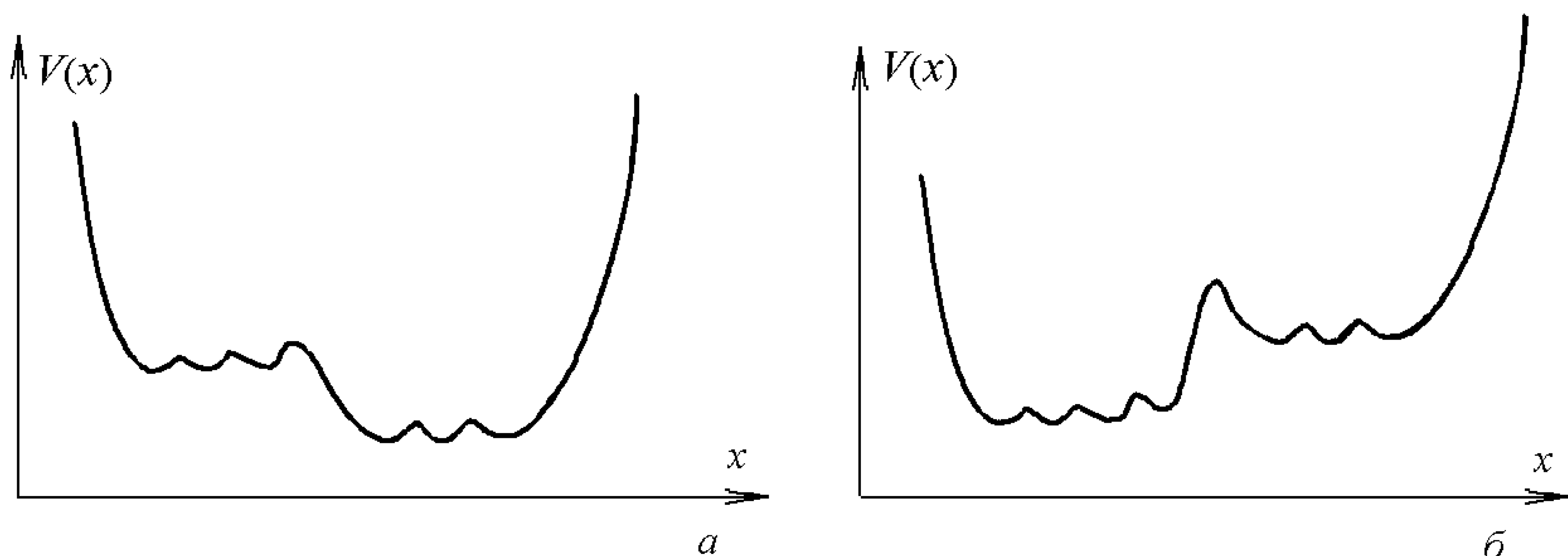


Рис. 2.23. Графики потенциала, определяющего состояние климата: в межледниковый период (*а*); в ледниковый период (*б*)

колебания температур. Прецессия — колебания земной оси вокруг полюса эклиптики — носит весьма сложный характер и спектр ее колебаний распадается на две области с периодами примерно 19 и 23 тыс. лет. Совместное влияние всех трех факторов способно изменять среднегодовую температуру относительно средней многолетней на 5%.

Теория Миланковича согласуется с геологическими данными о периодах оледенения. Однако подробное численное моделирование, включающее влияние параметров орбиты, процессы тепло- и массопереноса в атмосфере и океане Земли требует решения сложных нелинейных дифференциальных уравнений в частных производных с разрывными, вообще говоря, коэффициентами. Поэтому рассмотрим упрощенный подход, включающий в рассмотрение единственную переменную, например, среднегодовую температуру в какой-нибудь точке Земного шара, и предположим, что эволюция этого состояния определяется уравнением типа (2.18) с потенциалом, график которого изображен на рис. 2.23. Пусть внешние параметры таковы, что график потенциала имеет вид, приведенный на рис.2.23,*a*. Малые возмущения системы могут перевести систему из одного устойчивого состояния локального минимума потенциала в другое, однако чтобы перевести систему в состояние с малым значением координаты x (соответствующим среднегодовой температуре, характерной для ледникового периода), требуется невероятно большое возмущение. Однако при изменении орбиты Земли график потенциала может измениться от изображенного слева к изображенному на рис. 2.23,*б*, и тогда малые возмущения системы могут достаточно быстро вытолкнуть ее в состояние ледникового периода.

Если не интересоваться деталями изменения климата, соответствующими локальным минимумам, то можно заключить, что модальность и быстрые переходы от одного устойчивого состояния к другому свидетельствуют о возможности катастрофы типа “сборка”.

Глава 3

Приближенные методы исследования нелинейных систем

Как уже отмечалось, получить точное решение системы дифференциальных уравнений, описывающих динамику нелинейных систем, можно только в редких случаях. Поэтому на практике широко распространены методы приближенных исследований, в которых вместо заданной системы, не поддающейся точному анализу, изучается система, в каком-то смысле близкая к заданной, допускающей точное исследование. В этом разделе рассматриваются способы количественного изучения систем, основанные на методах малого параметра, а также метод усреднения.

3.1. Метод усреднения

Пусть эволюция системы характеризуется двумя временными масштабами — например, быстрыми осцилляциями и медленным дрейфом, как показано на рис. 3.1. Иногда для того, чтобы получить первоначальное представление о поведении системы, достаточно знать ее поведение "в целом", без мелкомасштабных временных подробностей. С этой целью можно усреднить эволюционные уравнения по "быстрому" времени и получить новое описание системы, характеризующее ее "среднюю" эволюцию. Малый параметр здесь — отношение масштабов "быстрого" и "медленного" времени. Простейшие механические системы, в которых материальная точка движется под действием внешних сил, как правило, поддаются формальному анализу, в общем же случае важную роль играет физическая интуиция.

Формальный подход проиллюстрируем на следующем примере.

Если в системе есть быстрые осцилляции, то фазовый портрет системы будет выглядеть как спираль, см.рис. 3.1, и движение системы можно описать в полярных координатах $x_1 = I\cos\vartheta$, $x_2 = I\sin\vartheta$.

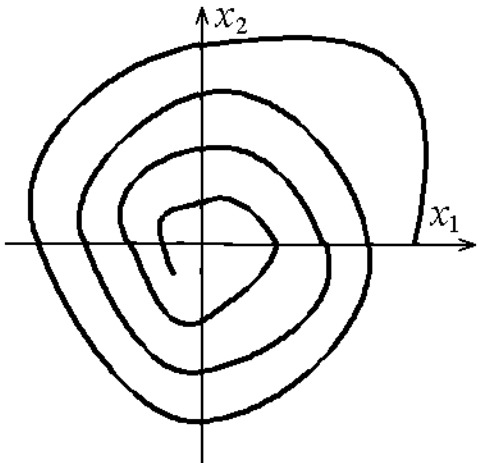


Рис. 3.1. Фазовая траектория системы, исследуемой методом усреднения

В примере с линейным осциллятором, см. раздел 1.2.1, роль переменной I играет полная энергия E системы. В механических системах это канонические переменные "действие–угол". Анализируя фазовый портрет, легко заметить, что переменная ϑ с течением времени изменяется значительно быстрее, чем переменная I. Усредняя по углу (быстрая переменная), можно получить уравнения для описания только одной медленной переменной I.

Пусть система описывается соотношениями

$$\dot{I} = \varepsilon f(I,\vartheta), \quad \dot{\vartheta} = \omega(t) + \varepsilon g(I,\vartheta).$$

Усредняя по фазе ϑ, имеем уравнение

$$\dot{\overline{I}} = \varepsilon \overline{f}(\overline{I}) \tag{3.1}$$

с начальным условием $\overline{I}(0) = I(0)$. Решение (3.1) сводится к квадратурам.

В качестве другого примера рассмотрим встречавшееся нам уже раньше уравнение Ван дер Поля, см. раздел 1.2.6:

$$\ddot{x} + \gamma\dot{x}(x^2/a_0^2 - 1) + \omega^2 x = 0. \tag{3.2}$$

В невозмущенном случае $\gamma = 0$ при должным образом заданных начальных условиях оно описывает линейный осциллятор. Введем полярные координаты (I,ϑ), в которых невозмущенная фазовая траектория представляет собой окружность:

$$x = \left(\frac{2I}{\omega}\right)^{1/2}\cos\vartheta, \quad \dot{x} = (2I\omega)^{1/2}\sin\vartheta. \tag{3.3}$$

Умножая обе части уравнения (3.2) на $\dot{x}/\omega$, $\omega > 0$, получим

$$\frac{\dot{x}}{\omega}\frac{d}{dt}\dot{x} + \frac{\gamma}{\omega a_0^2}\dot{x}^2(x^2 - a_0^2) + \omega\dot{x}x = \frac{d}{dt}\left(\frac{\dot{x}^2 + \omega^2 x^2}{2\omega}\right) +$$

$$+ \frac{\gamma}{\omega a_0^2}\dot{x}^2\left(x^2 - a_0^2\right) = 0.$$

Далее, подставляя (3.3) в уравнение (3.2) и интегрируя по периоду, меняя порядок дифференцирования и интегрирования, получим соотношение для усредненной по периоду переменной $\overline{I}$:

$$\dot{\overline{I}} = -\varepsilon\omega\frac{\overline{I}}{I_0}\left(\overline{I} - I_0\right), \tag{3.4}$$

где $\varepsilon = \text{const}$, $I_0 = \text{const}$. Решение уравнения (3.4) описывает систему с двумя положениями равновесия: $\overline{I}_1 = 0$ и $\overline{I}_2 = I_0$. В окрестности $\overline{I}_1$ выполнено $\dot{\overline{I}} \approx \varepsilon\omega\overline{I}$, и точка $\overline{I} = 0$ — точка неустойчивого равновесия, так как в ее окрестности $\overline{I}$ растет (при $\varepsilon > 0$). $\overline{I}_1$ является неустойчивым фокусом. В окрестности $\overline{I}_2$ $\dot{\overline{I}} \approx -\varepsilon\omega(\overline{I} - I_0)$, и движение устойчиво — это устойчивый предельный цикл.

Если $\varepsilon < 0$, то — наоборот, при $\overline{I} = 0$ имеем устойчивый фокус, а при $\overline{I} = I_2$ — неустойчивый предельный цикл.

3.2. Асимптотические методы малого параметра

В рассматриваемых в этом разделе методах подразумевается, что при значении малого параметра, равного нулю, мы получаем модель динамической системы в виде начальной задачи для системы уравнений, точное решение которой известно. При значениях параметра, отличных от нуля, система называется возмущенной. Точного решения начальной задачи в этом случае получить не удается, однако можно попытаться получить ее приближенное решение, воспользовавшись малостью параметра. Методы называются асимптотическими, так как при стремлении параметра к нулю приближенное решение стремится к точному решению невозмущенной задачи. Решение, получаемое асимптотическими методами при фиксированном и отличном от нуля параметре, обычно представляет собой полином некоторой степени от параметра, причем эту степень можно увеличивать сколь угодно, однако в результате мы рискуем получить ряд, который, вообще говоря, не будет сходится, поэтому на практике в качестве приближенного решения имеет смысл использовать лишь небольшое число членов этого ряда.

Различают сингулярные и регулярные возмущения. При сингулярных порядок системы при отличном от нуля значении параметра больше, чем порядок невозмущенной системы. К регулярным отнесем такие возмущения, которые не меняют порядок системы дифференциальных уравнений.

3.2.1. Регулярные возмущения системы дифференциальных уравнений.

Представление решения в виде ряда по степеням малого параметра. Пусть уравнение, описывающее эволюцию системы, содержит малый параметр. Наиболее распространенные методы исследования таких систем состоят в разложении решения по этому параметру; т.е. представлению решения в виде полинома, зависящего от малого параметра, и поиске коэффициентов этого полинома. Чем выше степень полинома, тем меньше вклад старших степеней при малых значениях параметра. Заметим, однако, что "лобовой" принцип поиска решения в виде полинома не всегда приводит к успеху. В доказательство этого утверждения рассмотрим следующий пример.

Пусть эволюционное уравнение задано в виде

$$\dot{x} = -\varepsilon x.$$

Точное его решение имеет вид $x = Ce^{-\varepsilon t}$. Но если искать решение в виде ряда по малому параметру

$$x = a_0 x_0(t) + \varepsilon x_1(t) + \ldots,$$

то получим, что

$$x = C(1 - \varepsilon t + \varepsilon^2 t^2/2 + \ldots)$$

Оборвав ряд, заметим, что приближение годится лишь для $t \ll 1/\varepsilon$, т.е. для времени, за которое x не успевает заметно измениться.

Метод Пуассона для уравнения колебаний. Этот пример также демонстрирует трудности, возникающие на пути "лобового" применения метода. Рассмотрим уравнение

$$\ddot{x} + \omega^2 x = \varepsilon f(x, \dot{x}) \tag{3.5}$$

и будем искать его приближенное решение, удовлетворяющее (3.5) с точностью до величин порядка ε^{n+1}, в виде n-й частичной суммы ряда

$$x(t) = x_0(t) + \varepsilon x_1(t) + \ldots + \varepsilon^n x_n(t) + \ldots \tag{3.6}$$

Подставляя (3.6) в (3.5), получим систему

$$\begin{aligned}
\ddot{x}_0 + \omega^2 x_0 &= 0,\\
\ddot{x}_1 + \omega^2 x_1 &= f(x_0, \dot{x}_0),\\
\ddot{x}_2 + \omega^2 x_2 &= f(x_1, \dot{x}_1),\\
&\ldots\ldots\ldots\ldots
\end{aligned}$$

Однако при решении ее возникают так называемые секулярные члены, т.е. величины вида $t^n \sin t$ и $t^n \cos t$. Действительно, пусть в (3.5) $f(x, \dot{x}) = x^3$, так, что

$$\ddot{x} + \omega^2 x + \varepsilon x^3 = 0, \tag{3.7}$$

и в безразмерных единицах $0 < \varepsilon \ll \omega^2$. Таким образом, уравнение (3.5) описывает колебания нелинейного осциллятора с возвращающей силой, близкой по своим характеристикам к линейной. С качественной точки зрения уравнение (3.7) описывает движение в потенциальном поле сил с потенциалом вида $\dfrac{\varepsilon x^4}{4} + \dfrac{\omega^2 x^2}{2}$; при $\varepsilon \geq 0$ имеется одно устойчивое состояние равновесия $x = 0$, и эволюция системы топологически эквивалентна гармоническим колебаниям.

Рассмотрим, к чему приводит метод Пуассона. Ограничившись первой степенью ε в решении (3.6), имеем

$$x = x_0 + \varepsilon x_1,$$

где

$$\ddot{x}_0 + \omega^2 x_0 = 0, \quad \ddot{x}_1 + \omega^2 x_1 = -x_0^3.$$

Из первого уравнения $x_0 = a\cos(\omega t + \varphi)$, и для второго уравнения получаем резонанс:

$$\ddot{x}_1 + \omega^2 x_1 = -\frac{3}{4}a^3\cos(\omega t + \varphi) - \frac{1}{4}a^3\cos 3(\omega t + \varphi),$$

так, что амплитуда колебаний растет линейно со временем:

$$x = a\cos(\omega t + \varphi) - \frac{3\varepsilon}{8\omega}a^3 t\sin(\omega t + \varphi) + \frac{\varepsilon a^3}{32\omega^2}\cos 3(\omega t + \varphi),$$

и при больших t даже качественно не имеет ничего общего с точным решением, выражающимся через эллиптические функции; система, описываемая моделью (3.5) с кубической правой частью, совершает колебательное движение с потенциалом $U(x) = \omega^2 x^2/2 + \varepsilon x^4/4$, близким к квадратичному.

Заметим, что причина такого поведения приближенного решения состоит в неудачном представлении решения. Действительно, запишем через разложение в ряд функцию $\sin(\omega + \varepsilon)t$ при малых ε. Для любых ω и t

$$\sin(\omega + \varepsilon)t = \sin\omega t + \varepsilon t\cos\omega t - \frac{\varepsilon^2 t^2}{2!}\sin\omega t + o(\varepsilon^2 t^2)$$

и трудно установить периодическое поведение разлагаемой функции.

Метод Крылова–Боголюбова. Рассмотрим построение асимптотического решения в системах, близких к линейным колебательным, т.е. описываемым уравнением вида

$$\ddot{x} + \omega^2 x = \varepsilon f(x, \dot{x}), \quad 0 < \varepsilon \ll 1. \tag{3.8}$$

При $\varepsilon = 0$ имеем случай линейных колебаний:

$$x = a\cos\psi, \quad \psi = \omega t + \vartheta, \tag{3.9}$$

где константы a и ϑ определяются начальными условиями.

Из физических соображений ясно, что наличие нелинейности приводит в (3.9) к зависимости a и ψ от времени t так, что $\dot{a} \neq 0$, $\dot{\psi} \neq \text{const}$. Будем искать решение в виде

$$x = a\cos\psi + \varepsilon u_1(a,\psi) + \varepsilon^2 u_2(a,\psi) + \ldots, \tag{3.10}$$

где u_1, $u_2 \ldots$ — периодические функции угла ψ с периодом 2π, а a и ψ как функции времени удовлетворяют дифференциальным уравнениям

$$\begin{gathered} \dot{a} = \varepsilon A_1(a) + \varepsilon^2 A_2(a) + \ldots, \\ \dot{\psi} = \omega + \varepsilon B_1(a) + \varepsilon^2 B_2(a) + \ldots \end{gathered} \tag{3.11}$$

Задача состоит в выборе выражений для функций $u_i(a,\psi)$, $A_i(a)$, $B_i(a)$, $i = 1, 2, \ldots$, так, чтобы (3.10), будучи подставлено в (3.8), удовлетворяло ему. Но на практике из-за возрастания сложности вычислений можно найти решения для второй – третьей степени ε в (3.10), поэтому для нас не важна сходимость суммы (3.10), а ее асимптотические свойства при $\varepsilon \to 0$ и фиксированном числе слагаемых. То есть при достаточно малых ε выражение (3.10) с конечным числом слагаемых должно достаточно точно представлять решение задачи (3.8) для достаточно длительных интервалов времени. Это — так называемый асимптотический метод построения приближенного решения (3.8). Более подробно он изложен в работе [3].

Иными словами, построение асимптотического решения состоит в том, чтобы выбрать функции

$$u_i(a,\psi), \quad A_i(a), \quad B_i(a), \quad i = 1, 2, \ldots, m, \tag{3.12}$$

так, чтобы выражение (3.10) с $(m+1)$-м слагаемым удовлетворяло уравнению (3.8) с точностью до величин порядка ε^{m+1} (так называемому "уравнению m-го приближения").

Замечание. Если уравнение (3.5) удовлетворяется с точностью до величин порядка ε^{m+1}, то, согласно общей теории обыкновенных дифференциальных уравнений, отличие приближенного решения от точного (при согласовании начальных условий) ограничивается величиной порядка $\varepsilon^{m+1}t$ (а не εt, как в методе Пуассона).

Замечание. Представление (3.10), (3.11) для функций (3.12) неоднозначно. Действительно, функции a и ψ в (3.10) могут быть определены с точностью до слагаемых порядка ε. Поэтому наложим дополнительные условия на функции u_i, $i = 1, 2, \ldots$, состоящие в том,

чтобы нулевое приближение описывало полную амплитуду первой гармоники колебаний, математически это выражается ортогональностью функций u_i, $i = 1, 2, \ldots$, к первой гармонике:

$$\int_0^{2\pi} u_i(a,\psi)\cos\psi\, d\psi = 0,$$
$$\int_0^{2\pi} u_i(a,\psi)\sin\psi\, d\psi = 0, \quad i = 1, 2, \ldots \tag{3.13}$$

Подставим (3.10) и (3.11) в исходное уравнение (3.8) и, представляя его правую часть по формуле Тейлора в точке $x = a\cos\psi$, $\dot{x} = -a\omega\sin\psi$, приравняем коэффициенты при одинаковых степенях ε. В результате получим систему дифференциальных уравнений для определения функций A_i, B_i, u_i, $i = 1, 2, \ldots$ Для упрощения записи обозначим

$$f_0(a,\psi) = f(a\cos\psi,\ -a\omega\sin\psi),$$

$$\begin{gathered} f_1(a,\psi) = u_1 f'_x(a\cos\psi,\ -a\omega\sin\psi) + \\ +[A_1\cos\psi - aB_1\sin\psi + +\omega(u_1)'_x] f'_{x'}(a\cos\psi,\ -a\omega\sin\psi) + \\ +(aB_1^2 - A_1A_1')\cos\psi + (2A_1B_1 + A_2B_1'a)\sin\psi - 2\omega A_1(u_1)''_{a\psi} - \\ -2\omega B_1(u_1)''_{\psi\psi}, \end{gathered}$$

и т.д., тогда

$$\begin{aligned} \omega^2((u_1)''_{\psi\psi} + u_1) &= f_0(a,\psi) + 2\omega A_1\sin\psi - 2\omega aB_1\cos\psi, \\ \omega^2((u_2)''_{\psi\psi} + u_2) &= f_1(a,\psi) + 2\omega A_2\sin\psi - 2\omega aB_2\cos\psi, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ \omega^2((u_n)''_{\psi\psi} + u_2) &= f_n(a,\psi) + 2\omega A_n\sin\psi - 2\omega aB_n\cos\psi, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned} \tag{3.14}$$

Для решения системы (3.14) воспользуемся разложениями Фурье. Согласно обозначениям, $f_k(a,\psi)$ периодическая по ψ функция с периодом 2π; она определяется функциями A_i, B_i, u_i, $i = 1, 2, \ldots$ Чтобы найти функции u_1, A_1, B_1, запишем разложения Фурье для функций f_0 и u_1:

$$\begin{aligned} f_0(a,\psi) &= g_0(a) + \textstyle\sum_{n=1}^{\infty}\{g_n(a)\cos n\psi + h_n(a)\sin n\psi\}, \\ u_1(a,\psi) &= v_0(a) + \textstyle\sum_{n=1}^{\infty}\{v_n(a)\cos n\psi + w_n(a)\sin n\psi\}, \end{aligned}$$

Подставим это в (3.14) и дифференцируя под знаком ряда, получим

$$\begin{gathered} \omega^2 v_0(a) + \textstyle\sum_{n=1}^{\infty}\omega^2(1-n^2)\{v_n(a)\cos n\psi + w_n(a)\sin n\psi\} = \\ = g_0(a) + \{g_1(a) + 2\omega aB_1\}\cos\psi + \{h_1(a) + 2\omega A_1\}\sin\psi + \\ + \textstyle\sum_{n=2}^{\infty}\{g_n(a)\cos n\psi + h_n(a)\sin n\psi\}, \end{gathered} \tag{3.15}$$

откуда, приравнивая коэффициенты при одинаковых гармониках, найдем A_1, $\mathcal{B}_1$, v_0, v_i, w_i, $i = 2, 3, \ldots$ При $n = 1$ коэффициенты при v_1 и w_1 не определяются, однако из условия ортогональности функции u_1 основной гармонике следует, что $v_1 = w_1 = 0$. Окончательно

$$u_1(a,\psi) = \frac{g_0(a)}{\omega^2} + \omega^{-2}\sum_{n=2}^{\infty}\left(\frac{1}{1-n^2}\right)\{g_n(a)\cos n\psi + h_n(a)\sin n\psi\},$$

$$A_1 = -\frac{h_1(a)}{2\omega}, \quad B_1 = -\frac{g_1(a)}{2\omega a}.$$

Теперь мы имеем все для определения явного выражения для $f_1(a,\psi)$; повторяя эту процедуру и разлагая в ряд Фурье $f_1(a,\psi)$, получим функцию u_2 и т.д.

Итак, рассмотрены принципы построения асимптотического решения для динамических систем и получено выражение для первого приближения. Запишем его в виде

$$x = a\cos\psi + \varepsilon u_1(a,\psi), \tag{3.16}$$

где

$$\dot{a} = \varepsilon A_1(a), \quad \dot{\psi} = \omega + \varepsilon B_1(a),$$

т.е.

$$\Delta a = a(t) - a(0) \sim \varepsilon t\widetilde{A}_1, \quad \Delta(\psi - \omega t) \sim \varepsilon t\widetilde{B}_1;$$

здесь $\widetilde{A}_1$ и $\widetilde{B}_1$ — некоторые средние значения $A_1(t)$ и $B_1(t)$ на интервале $(0,t)$. Это означает, что заметные приращения a и ψ могут получить на интервалах времени порядка $1/\varepsilon$. Но уравнения первого приближения получены из уравнений (3.11) отбрасыванием членов порядка ε^2, значит, сами функции a и ψ на интервале $(0,t)$ определены с точностью $\varepsilon^2 t$. Учитывая это и исходя из (1), точное решение $\widehat{x}(t)$ исходного уравнения можно записать в виде

$$\widehat{x}(t) = \left[a(t)\cos\psi(t) + O(\varepsilon^2 t) + \varepsilon u_1(a,\psi)\right]|_{t\sim 1/\varepsilon} = \\ = a\cos\psi + O(\varepsilon) + \varepsilon u_1(a,\psi),$$

таким образом не имеет смысла сохранять слагаемое $\varepsilon u_1(a,\psi)$ в формуле первого приближения (1).

Нелинейное трение. Рассмотрим движение материальной точки, описываемое уравнением

$$\ddot{x} + \omega^2 x = \varepsilon F(\dot{x}).$$

с некоторыми начальными условиями. Это уравнение описывает колебание единичной массы под действием упругой силы и нелинейного

слабого трения, зависящего от скорости. Оно приводится к рассмотренному выше уравнению, так как

$$f(x, \dot{x}) = F(\dot{x}).$$

Для того чтобы воспользоваться формулами первого приближения, разложим в ряд Фурье функцию $f_0(a, \psi) = f(a\cos\psi, -a\omega\sin\psi) = F(-a\omega\sin\psi)$. Так как $F(a\cos\psi)$ как функция ψ является четной, то

$$F(a\cos\psi) = \sum_{n=0}^{\infty} F_n(a)\cos n\psi,$$

но

$$F(-a\omega\sin\psi) = F\left(a\omega\cos\left(\psi + \frac{\pi}{2}\right)\right) = \sum_{n=0}^{\infty} F_n(a\omega)\cos n(\psi + \frac{\pi}{2}) =$$

$$= \sum_{n=0}^{\infty} F_n(a\omega)\left[\cos n\psi\cos\frac{n\pi}{2} - \sin n\psi\sin\frac{n\pi}{2}\right],$$

а значит,

$$g_n(a) = F_n(a\omega)\cos\frac{n\pi}{2}, \quad h_n(a) = -F_n(a\omega)\sin\frac{n\pi}{2},$$

откуда

$$A_1(a) = \frac{1}{2\omega}F_1(a\omega), \quad B_1(a) = 0,$$

и первое приближение задается соотношением

$$x = a\cos\psi, \quad \dot{a} = \frac{\varepsilon}{2\omega}F_1(a\omega), \quad \dot{\psi} = \omega,$$

т.е. в первом приближении движение системы описываются затухающими (или возрастающими, в зависимости от знака ε) гармоническими колебаниями с постоянной частотой.

3.2.2. Асимптотические решения сингулярно возмущенных дифференциальных уравнений. Динамические системы, поведение которых определяется несколькими процессами с различными временными масштабами, часто встречаются в приложениях. Примером могут служить процессы в химических реакторах, где, наряду с достаточно медленными химическими реакциями, идут и быстрые; в математических моделях развития биологических популяций часто приходится учитывать и популяции с периодом жизни в несколько лет (птицы, млекопитающие), и с жизненным периодом в несколько дней

(насекомые). При описании таких процессов возникают дифференциальные уравнения, в которых малый параметр стоит при старшей производной — это так называемые сингулярно возмущенные уравнения. Первые работы, посвященные математическим методам исследования и решения таких уравнений, принадлежат А.Н.Тихонову.

Простейшие сингулярно возмущенные задачи. Рассмотрим, в чем состоят основные проблемы теории сингулярных возмущений, и опишем идею методов решения сингулярно возмущенных уравнений, на следующем простом примере. Рассмотрим начальную задачу для дифференциального уравнения

$$\mu\frac{dx}{dt} = F(x,t) \tag{3.17}$$

с начальным условием

$$x|_{t=t_0} = x_0, \tag{3.18}$$

где μ — малый параметр. Невозмущенная задача — это задача, в которой в (3.17) положено $\mu = 0$, и при “обычном”, не сингулярном, возмущении такое решение, как правило, приближает решение возмущенной задачи (3.17), (3.18). Однако в нашем примере при $\mu = 0$ мы имеем систему алгебраических соотношений

$$F(x,t) = 0, \tag{3.19}$$

$$x|_{t-t_0} = x_0, \tag{3.20}$$

которая может оказаться несовместной. Действительно, если (3.19) разрешимо относительно x, и $x = \varphi(t)$ — одно из его решений, то оно вовсе не обязательно удовлетворяет дополнительному начальному условию (3.20), состоящему в выполнении при $t = t_0$ равенства $\varphi(t_0) = x_0$.

Таким образом, если для некоторого решения $\varphi(t)$ уравнения (3.19) выполнено равенство $\varphi(t_0) = x_0$, то решения возмущенной и невозмущенной задач совпадают. Если же это соотношение не выполнено ни для одного решения уравнения (3.19), то в точке $t = t_0$ при любом $\mu \neq 0$ решение возмущенной задачи (3.17), (3.18) не равно ни одному из решений невозмущенной задачи (3.19), и не стремится к нему при $\mu \to 0$, то есть в точке $t = t_0$ решение невозмущенной задачи не аппроксимирует решение точной (возмущенной). Однако возникает вопрос: раз в окрестности начальной точки нет аппроксимации, может, она возникнет позже, при достаточно больших t? Действительно, такое возможно. Пусть, например, имеется три решения алгебраического уравнения (3.19) — $\varphi_1(t)$, $\varphi(t)$ и $\varphi_2(t)$, графики которых приведены на рис. 3.2. Если правая часть уравнения (3.19) имеет знаки, указанные на рис. 3.2, то с ростом t решение $x(t)$ возмущенной задачи будет стремится к корню $\varphi(t)$ правой

части (3.17), если начальное условие x_0 удовлетворяет неравенству $\varphi_1(t_0) < x_0 < \varphi_2(t_0)$, так как если в некоторый момент времени $t > t_0$ выполняется неравенство $x(t) < \varphi(t)$, то $F(x,t) > 0$ и согласно уравнению (3.17), $x(t)$ возрастает, т.е. стремится к $\varphi(t)$ с ростом t; если же в этот момент времени выполняется противоположное неравенство $x(t) > \varphi(t)$, то $F(x,t) < 0$ и $x(t)$ с ростом t убывает к $\varphi(t)$. Таким образом, в рассматриваемом случае при $t \to \infty$ решение начальной задачи (3.17), (3.18) при любом $\mu > 0$ стремится к $\varphi(t)$ — решению невозмущенного уравнения (3.19). Очевидно, такое приближение не является равномерным; для того, чтобы при $\mu \to 0$ получить приближение решения возмущенной задачи, равномерное по всем $t > t_0$, необходимо прибавить к функции $\varphi(t)$ некоторую функцию, "исправляющую" ее поведение в окрестности начального значения $t = t_0$; такое слагаемое носит название пограничной функции.

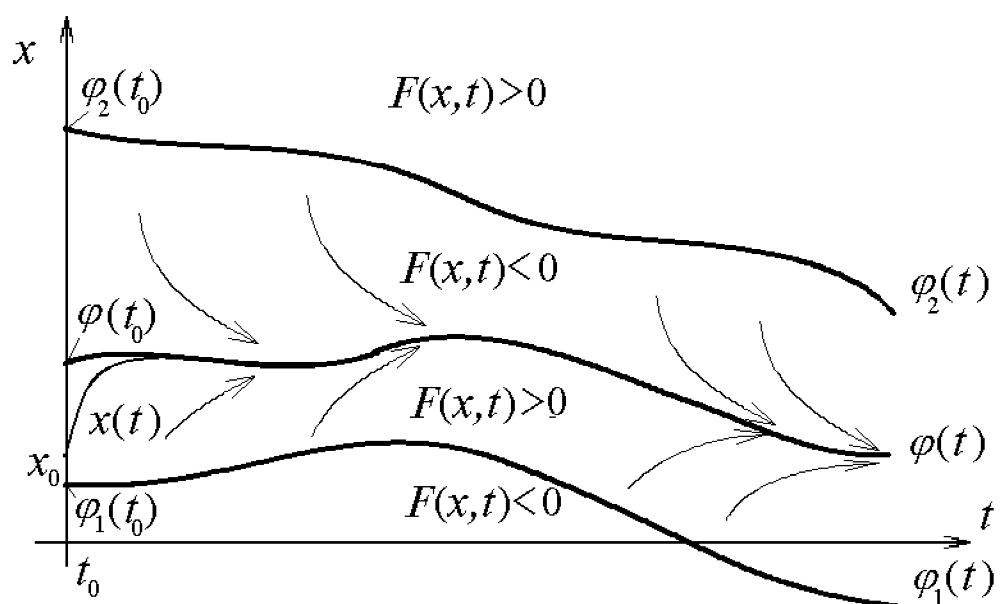


Рис. 3.2. Выбор приближенного решения скалярной сингулярно возмущенной задачи

Заметим также, что в рассматриваемом случае при начальном условии, удовлетворяющем неравенству $x_0 < \varphi_1(t_0)$, ни одно из решений уравнения (3.19) не является приближением решения возмущенной задачи.

Теорема Тихонова. Рассмотрим подробнее особенности сингулярно возмущенных задач на несколько более общем примере, имеющим непосредственное отношение к теории динамических систем.

Запишем систему уравнений, описывающих поведение динамической системы, в которой малый параметр μ является множителем при производной только в первом уравнении:

$$\begin{aligned} \mu\dot{z} &= F(z,y,t), \\ \dot{y} &= f(z,y,t), \quad 0 \le t \le T, \end{aligned} \tag{3.21}$$

с начальным условием

$$z(0) = z_0, \quad y(0) = y_0. \tag{3.22}$$

Положив $\mu = 0$, придем к невозмущенной системе

$$\begin{aligned} 0 &= F(\overline{z},\overline{y},t), \\ \dot{\overline{y}} &= f(\overline{z},\overline{y},t), \end{aligned} \tag{3.23}$$

для которой достаточно задать лишь одно начальное условие

$$\overline{y}(0) = y_0. \tag{3.24}$$

Возникают вопросы: во-первых, как решать невозмущенную систему (3.23), (3.24). Из первого уравнения (3.23) выразим $\overline{z} = \varphi(\overline{y}, t)$. Заметим, что может оказаться, что это возможно не единственным образом, тогда возникнет вопрос о выборе корня. Пусть корень выбран, тогда, подставив его в (3.23) в уравнение для $\overline{y}(t)$, найдем решение (3.23), (3.24) $\overline{y} = \overline{y}(t)$, $\overline{z} = \varphi(\overline{y}, t)$. Как оно соотносится с (3.21), (3.22)? Ясно, что, вообще говоря, оно не удовлетворяет (3.22), так как начальное условие $z(0) = z_0$ не участвовало в построении решения $\overline{y} = \overline{y}(t)$, $\overline{z} = \varphi(\overline{y}, t)$, значит, последнее не будет приближать решение (3.21), (3.22) и в окрестности точки $t = 0$. Но может быть, оно будет приближать решение (3.21), (3.22) вне этой окрестности?

В отличие от z, $\overline{y}$ близко к y_0 в окрестности $t = 0$. Сохранится ли эта близость дальше? Ответы на эти вопросы содержатся в теореме Тихонова, к обсуждению которой мы переходим.

Сформулируем три условия.

I. *Пусть уравнение* $F(\overline{z}, \overline{y}, t) = 0$ *имеет изолированный корень относительно* $\overline{z}$: $\overline{z} = \varphi(\overline{y}, t)$, $(\overline{y}, t) \in D$, *где* D — *некоторая область, а задача*

$$\dot{\overline{y}} = f(\varphi(\overline{y}, t), \overline{y}, t), \quad \overline{y}(0) = y_0 \tag{3.25}$$

имеет единственное решение на отрезке $[0, T]$, *соответствующее этому корню.*

Наглядно корень $z = \varphi(y, t)$ можно представить как некоторую поверхность в пространстве (y, z, t), на которой функция F равна нулю, рис. 3.3. Изолированность корня значит, что в некоторой ее окрестности нет других корней.

Прежде, чем переходить к формулировке следующих двух условий, рассмотрим так называемую присоединенную систему уравнений. Обозначим через $\tau = t/\mu$ "растянутое" время, тогда (3.21), (3.22) примет вид

$$\begin{aligned} z'_\tau &= F(z, y, \tau\mu), \\ y'_\tau &= \mu f(z, y, \tau\mu). \end{aligned} \tag{3.26}$$

При малых μ правая часть (3.26) мала, и скорость изменения y по отношению к растянутому времени имеет порядок $O(\mu)$, поэтому $y = y_0 + O(\mu)$ при $0 \le \tau \le \tau_0$. Тогда

$$z'_\tau = F(z, y_0 + O(\mu), \tau\mu) \tag{3.27}$$

есть регулярно возмущенное по отношению к

$$z'_\tau = F(z, y_0, 0) \tag{3.28}$$

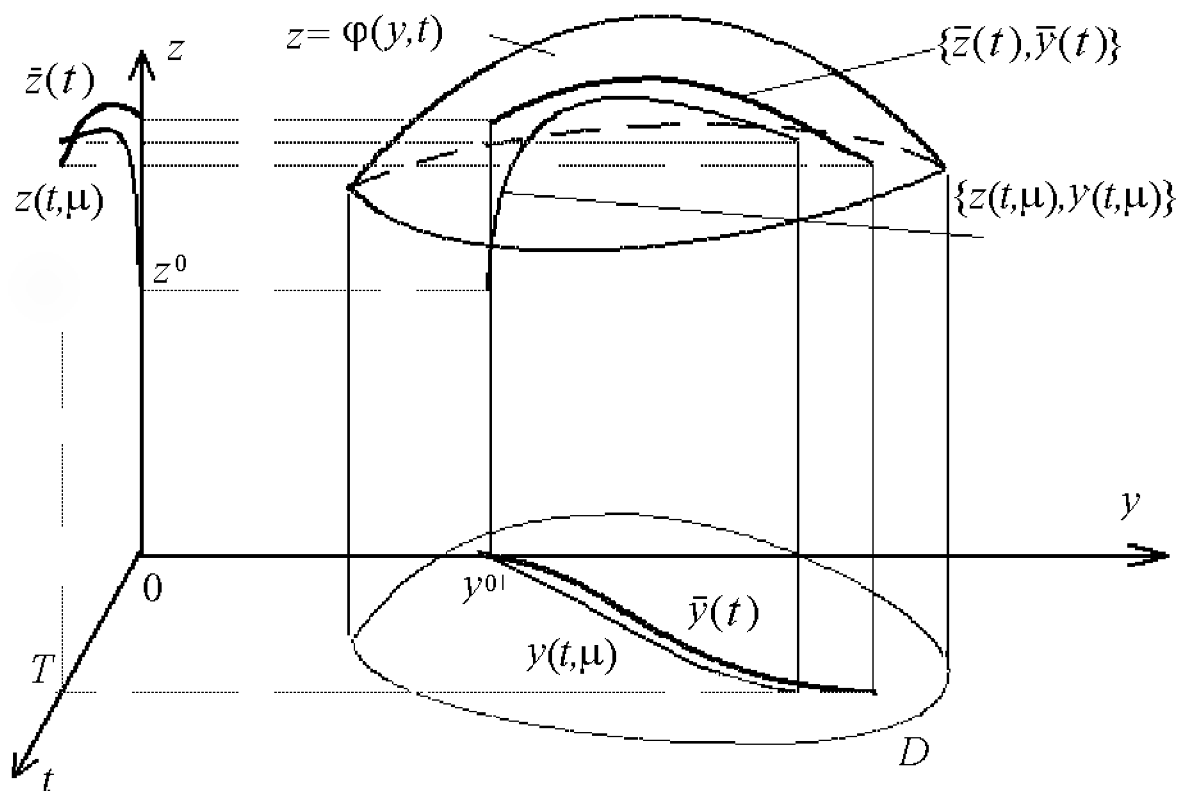


Рис. 3.3. Точное и приближенное решения сингулярно возмущенной задачи

уравнение, и есть надежда, что при малых μ решения (3.27) и (3.28) будут близки. Соотношение

$$\widetilde{z}'_\tau = F(\widetilde{z}, y, t) \tag{3.29}$$

называется присоединенным уравнением. В нем $\widetilde{z} = \widetilde{z}(\tau)$, $y = y(t)$, и τ выступает как независимая переменная, а t — как параметр (вместе с $y(t)$).

В силу условия I $\widetilde{z} = \varphi(y,t)$ является решением уравнения (3.29); действительно, оно не зависит от τ и обращает правую часть в ноль. Такое не зависящее от τ решение называется точкой покоя.

II. *Пусть точка покоя $\widetilde{z} = \varphi(y,t)$ присоединенного уравнения* (3.29) *является асимптотически устойчивой по Ляпунову при $\tau \to \infty$ равномерно относительно $(y,t) \in D$, т.е. для любого $\varepsilon > 0$ существует такое $\delta(\varepsilon) > 0$ (одно и то же для всех $(y,t) \in D$), что если $|\widetilde{z}(0) - \varphi(y,t)| < \delta$, то выполняются условия*

$$|\widetilde{z}(\tau) - \varphi(y,t)| < \varepsilon \quad \forall \tau > 0, \tag{3.30}$$

$$\widetilde{z}(\tau) \to \varphi(y,t), \quad \tau \to \infty. \tag{3.31}$$

Рассмотрим присоединенное уравнение (3.29). При $y = y_0$, $t = 0$, т.е. при начальных значениях параметров y и t, оно имеет вид

$$\widetilde{z}'_\tau = F(\widetilde{z}, y_0, 0), \quad \tau > 0, \tag{3.32}$$

с начальным условием

$$\widetilde{z}(0) = z_0, \tag{3.33}$$

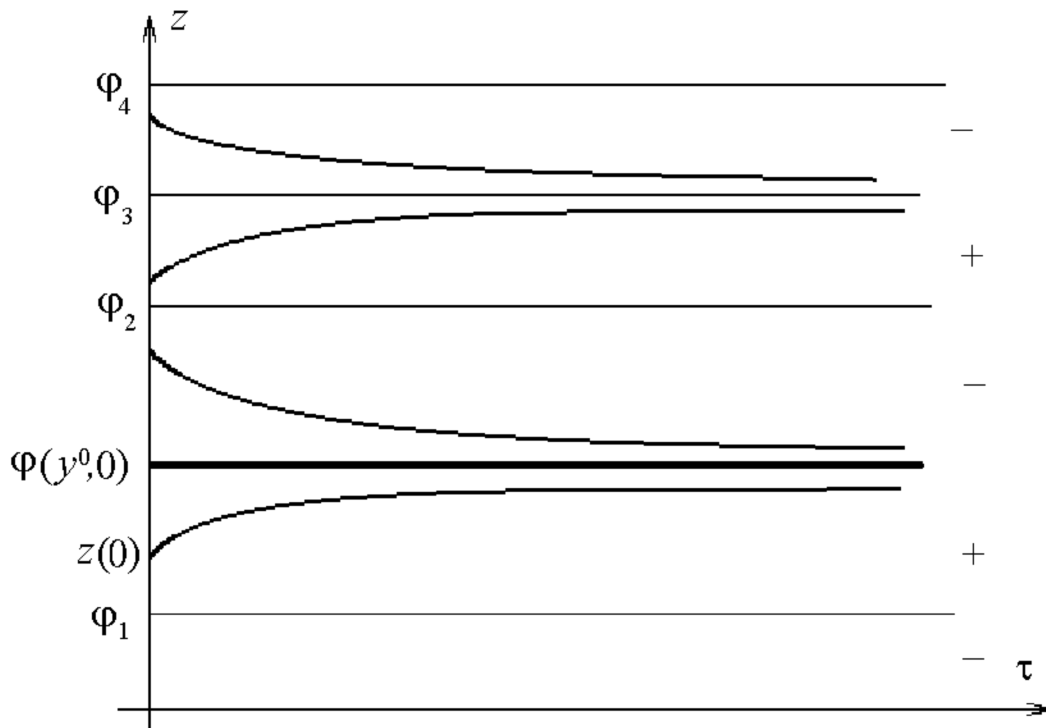


Рис. 3.4. Выбор приближенного решения сингулярно возмущенной задачи

т.е. с тем самым числом, которое фигурирует в исходном начальном условии (3.22). Это число, вообще говоря, может не быть близким к точке покоя $\varphi(y_0, 0)$, поэтому $\tilde{z}(\tau)$ может и не стремиться к точке покоя $\varphi(y, t)$ при $\tau \to \infty$. Потребуем этого в условии III.

III. *Пусть решение* $\tilde{z}(\tau)$ *задачи* (3.32), (3.33) *стремится к точке покоя* $\varphi(y_0, 0)$ *при* $\tau \to \infty$.

В таком случае говорят, что значение z_0 принадлежит области влияния точки покоя.

Сформулируем теперь теорему Тихонова [4].

Теорема. *Если выполнены условия* I–III, *то при достаточно малых* μ *задача* (3.21), (3.22) *имеет единственное решение* $z(t, \mu)$, $y(t, \mu)$, *и для этого решения справедливы предельные равенства*

$$\lim_{\mu \to 0} y(t, \mu) = \overline{y}(t), \quad 0 \le t \le T,$$

$$\lim_{\mu \to 0} z(t, \mu) = \overline{z}(t), \quad 0 < t \le T.$$

Теорема Тихонова гарантирует близость решения $\overline{z}(t)$, $\overline{y}(t)$ к точному $z(t, \mu)$, $y(t, \mu)$ при малых μ, иными словами, $\overline{y}(t)$ является асимптотическим приближением для $y(t, \mu)$ на отрезке $[0, T]$, а $\overline{z}(t)$ — асимптотическим приближением для $z(t, \mu)$ на отрезке $[\delta, T]$, $\delta > 0$.

Наиболее тонкий момент теоремы Тихонова — определение области влияния точки покоя. Для рассматриваемого скалярного случая может быть дана следующая наглядная интерпретация. Пусть уравнение $F(\tilde{z}, y_0, t)$ наряду с корнем $\tilde{z} = \varphi(y_0, 0)$ имеет и другие корни $\tilde{z}_i = \varphi_i$, $i = 1, 2, \ldots$ На плоскости $(\tilde{z}, \tau)$ эти корни изобразятся прямыми линиями, рис. 3.4.

Пусть в промежутках между корнями функция $F(\tilde{z}, y_0, 0)$ имеет такие знаки, как показаны на рис. 3.4. Так как $F(\tilde{z}, y_0, 0)$ равна производной z'_τ в силу (3.32), то если $\tilde{z}(0)$ лежит в промежутке от φ_1 до $\varphi(y_0, 0)$, т.е. там, где $F > 0$, то решение $\tilde{z}(\tau)$ при $\tau > 0$ будет возрастать и монотонно приближаться к точке покоя $\varphi(y_0, 0)$, если же $\tilde{z}(0)$ лежит между $\varphi(y_0, 0)$ и φ_2 т.е. там, где $F < 0$, то $\tilde{z}(\tau)$ при $\tau > 0$ будет убывать и снова приближаться к точке покоя $\varphi(y_0, 0)$. Если же взять $\tilde{z}(0)$ вне интервала (φ_1, φ_2), то $\tilde{z}(\tau)$ не будет стремиться к точке покоя $\varphi(y_0, 0)$.

Из рис. 3.4 видно, что точка φ_3 тоже является асимптотически устойчивой — с другой областью влияния (φ_2, φ_4). Поэтому условие II теоремы Тихонова не дает еще возможности однозначно выбрать нужный корень уравнения $F(z, y, t) = 0$, но условие III делает этот выбор однозначным. Заметим, что, вообще говоря, точка $\tilde{z}(0)$ может и не принадлежать области какой-либо асимптотически устойчивой точки покоя, тогда теорема Тихонова неприменима.

Теорема Тихонова для сингулярно возмущенных дифференциальных уравнений, таким образом, дает условия существования и единственности решения исходной сингулярно возмущенной задачи и условия построения приближенного асимптотического решения как решения исходной системы при малом параметре, равном нулю. Решения более высокой точности получим, если будем искать их в виде

$$\begin{aligned} z(t,\mu) &= \overline{z}_0(t) + \mu\overline{z}_1(t) + \ldots + \mu^k\overline{z}_k(t) + \ldots \\ &\ldots + \Pi_0 z(\tau) + \mu\Pi_1 z(\tau) + \ldots + \mu^k\Pi_k z(\tau) + \ldots, \\ y(t,\mu) &= \overline{y}_0(t) + \mu\overline{y}_1(t) + \ldots + \mu^k\overline{y}_k(t) + \ldots \\ &\ldots + \Pi_0 y(\tau) + \mu\Pi_1 y(\tau) + \ldots + \mu^k\Pi_k y(\tau) + \ldots, \end{aligned} \tag{3.34}$$

где $\tau = t/\mu$. Здесь $\Pi_k z(\tau)$, $\Pi_k y(\tau)$, $k = 0, \ldots,$ — так называемые пограничные функции, предназначенные для описания решения в окрестности начальной точки $t = 0$. Найти функции в (3.34) можно, подставляя (3.34) в исходное уравнение и, представляя его правые части по формуле Тейлора, приравнивая коэффициенты при одинаковых степенях μ. При этом придется наложить новые граничные условия $\Pi_k y(\infty) = 0$ (для функции $\Pi_k z(\cdot)$ аналогичные условия удовлетворятся автоматически, как следует из теоремы Тихонова).

Глава 4

Гамильтоновы системы

В теории нелинейных динамических систем важный класс составляют так называемые гамильтоновы системы, которые имеют наглядную механическую интерпретацию. Поскольку и лагранжева, и гамильтонова механика опираются на вариационные принципы, то прежде, чем переходить к рассмотрению таких систем, напомним ряд определений и понятий вариационного исчисления.

4.1. Основы вариационного исчисления

4.1.1. Основные понятия. Вариационное исчисление занимается отысканием экстремумов функций, определенных на бесконечномерных функциональных пространствах (пространствах кривых) со значениями на числовой прямой — такие функции называются функционалами. Для определенности будем считать множеством определения рассматриваемых здесь функционалов пространство $\mathcal{R}$ функций, свойства которых будем уточнять по мере надобности. Каждая такая функция является элементом (точкой) бесконечномерного линейного пространства.

Примером функционала является длина кривой на плоскости: $\gamma = \{(t,x) : x = x(t)\}$, где $x(\cdot)$ — дифференцируемая функция переменной $t \in [t_0, t_1] \subset \mathcal{R}_1$:

$$\Phi(\gamma) = \int\limits_{t_0}^{t_1} \sqrt{1+\dot{x}^2}\, dt.$$

Рассмотрим близкую к $\gamma \in \mathcal{R}$ кривую $\gamma' = \{(t,x) : x = x(t) + h(t)\}$,

$\gamma' \in \mathcal{R}$, обозначим ее $\gamma' = \gamma + h$ и рассмотрим приращение функционала $\Phi(\gamma + h) - \Phi(\gamma)$ (рис. 4.1).

Определение. Функционал $\Phi(\cdot)$ называется *дифференцируемым* в точке $\gamma \in \mathcal{R}$, если его приращение представимо в виде $\Phi(\gamma + h) - \Phi(\gamma) = F(\gamma, h) + R(\gamma, h)$, где $F(\cdot,\cdot)$ зависит от $h \in \mathcal{R}$ линейно, т.е. при фиксированном γ выполняется: $F(\gamma, h_1 + h_2) = F(\gamma, h_1) + F(\gamma, h_2)$, $F(\gamma, ch) = cF(\gamma, h)$, а $R(\gamma, h) = O(h^2)$, т.е. неравенства $| h | < \varepsilon$, $| dh/dt | < \varepsilon$ влекут $| R | < C\varepsilon^2$. Линейная по h часть $F(\cdot,\cdot)$ приращения функционала $\Phi(\cdot)$ называется его дифференциалом.

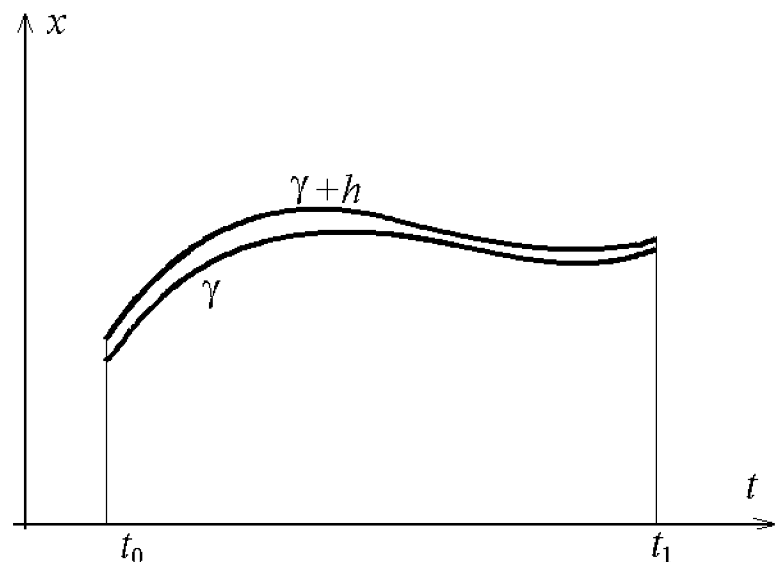


Рис. 4.1. Вариация функционала

Если $\Phi(\cdot)$ дифференцируем в каждой точке некоторого множества $\mathcal{M} \subset \mathcal{R}$, то он называется дифференцируемым на $\mathcal{M}$.

Если функционал дифференцируем, то его дифференциал определен однозначно. Он называется вариацией функционала $\Phi(\cdot)$, а функция (кривая) $h \in \mathcal{R}$ называется вариацией кривой γ.

4.1.2. Уравнение Эйлера–Лагранжа. Рассмотрим следующий пример. Пусть $\gamma = \{(t, x) : x = x(t), t_0 \le t \le t_1\}$ — кривая на плоскости (t, x), $\dot{x} = dx/dt$, $L = L(\cdot,\cdot,\cdot)$ — дифференцируемая функция трех переменных. Составим функционал

$$\Phi(\gamma) = \int_{t_0}^{t_1} L(x(t), \dot{x}(t), t)\, dt. \tag{4.1}$$

Тогда справедливо следующее утверждение.

Теорема. *Функционал* $\Phi(\cdot)$, *определенный в* (4.1), *дифференцируем, его дифференциал равен*

$$F(\gamma, h) = \int_{t_0}^{t_1} \left[\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}}\right] h\, dt + \left(\frac{\partial L}{\partial \dot{x}} h\right) \Big|_{t_0}^{t_1}.$$

Доказательство проводится применением формулы Тейлора и последующим интегрированием по частям.

Определение. *Экстремалью* дифференцируемого функционала $\Phi(\cdot)$ называется такая кривая $\gamma \in \mathcal{R}$, что $F(\gamma, h) = 0$ для всех h.

Теорема. *Для того чтобы* $\gamma = \{(t, x), x = x(t)\} \in \mathcal{R}$, *была экстремалью функционала* $\Phi(\gamma) = \int_{t_0}^{t_1} L(x(t), \dot{x}(t), t)\, dt$ *на пространстве*

кривых, проходящих через точки** $x(t_0) = x_0$, $x(t_1) = x_1$, **необходимо и достаточно, чтобы вдоль** $x(t)$ **выполнялось равенство

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = 0. \tag{4.2}$$

Определение. Уравнение (4.2) называется *уравнением Эйлера–Лагранжа* для функционала$\Phi(\gamma)$, определенного в (4.1).

Совершенно аналогично, если $\mathbf{x} \in \mathcal{R}_N$, $\gamma = \{(\mathbf{x}, t) \in \mathcal{R},\ t \in \mathcal{R}_1,\ \mathbf{x} \in \mathcal{R}_N,\ \mathbf{x} = \mathbf{x}(t), t_0 \le t \le t_1\}$ — кривая в $N+1$-мерном пространстве $\mathcal{R}_1 \times \mathcal{R}_N$, и $L : \mathcal{R}_1 \times \mathcal{R}_N \times \mathcal{R}_N \to \mathcal{R}_1$ — функция $2N+1$ аргументов, то для того, чтобы γ была экстремалью функционала

$$\Phi(\gamma) = \int\limits_{t_0}^{t_1} L(\mathbf{x}(t), \dot{\mathbf{x}}(t), t)\, dt$$

на пространстве кривых, соединяющих точки $(\mathbf{x}_0, t_0)$ и $(\mathbf{x}_1, t_1)$, необходимо и достаточно выполнения вдоль нее уравнения Эйлера–Лагранжа

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\mathbf{x}}} - \frac{\partial L}{\partial \mathbf{x}} = 0. \tag{4.3}$$

Это система из N уравнений 2-го порядка, ее решение зависит от $2N$ произвольных констант, для их определения надо задать $2N$ условий, например, $\mathbf{x}(t_0) = x_0$, $\mathbf{x}(t_1) = x_1$.

Заметим, что свойство кривой γ быть экстремалью функционала, не зависит от выбора системы координат. Так, если на плоскости задана декартова либо полярная система координат, то для длины кривой в разных системах даются разными формулами:

$$\Phi_{dec} = \int\limits_{t_0}^{t_1} \sqrt{\dot{x}_1^2 + \dot{x}_2^2}\, dt, \quad \Phi_{pol} = \int\limits_{t_0}^{t_1} \sqrt{\dot{r}^2 + r^2\dot{\varphi}^2}\, dt,$$

но экстремали в обеих случаях — одни и те же прямые линии.

4.2. Задачи динамики

Запишем второй закон Ньютона для системы материальных точек в потенциальном поле:

$$\frac{d}{dt}(m_i \dot{\mathbf{r}}_i) = -\frac{\partial U}{\partial \mathbf{r}_i}; \tag{4.4}$$

и уравнение Эйлера–Лагранжа имеет вид

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\mathbf{r}}_i} - \frac{\partial L}{\partial \mathbf{r}_i} = 0. \tag{4.5}$$

Как было показано в курсе теоретической механики, траектории движения механической системы (4.4) совпадают с экстремалями функционала $\overline{\Phi}(\gamma) = \int\limits_{t_0}^{t_1} L\,dt$, где $L = T - U$ — разность кинетической и потенциальной энергий. Действительно, $U = U(r)$, $T = \sum m_i \dot{r}_i^2$, $\partial L/\partial \mathbf{r}_i = -\partial U/\partial \mathbf{r}_i$. Значит, (4.5) — иная форма записи второго закона Ньютона.

Определение. В механике функцию $L = L(\mathbf{q}, \dot{\mathbf{q}}, t) = T - U$ называют *функцией Лагранжа* (*лагранжианом*), q_i — обобщенной координатой, $\dot{q}_i$ — обобщенной скоростью, $\partial L/\partial q_i$ — обобщенной силой, функционал $\overline{\Phi}(\gamma)$ — действием, а уравнение (4.3) — уравнением Лагранжа.

Заметим, что механическая интерпретация не обязательна, но весьма наглядна.

4.2.1. Движение в центральном поле. Циклические координаты. Рассмотрим движение точки в плоском центральном поле в полярных координатах $q_1 = r$, $q_2 = \varphi$. Кинетическая энергия равна $T = m(\dot{r}^2 + r^2\dot{\varphi}^2)/2$, и лагранжиан равен $L = T - U$, т.е. $L = T(q_1, q_2, \dot{q}_1, \dot{q}_2) - U(q_1) = m(\dot{q}_1^2 + q_1^2\dot{q}_2^2)/2 - U(q_1)$. Обобщенные импульсы даются формулами

$$p_1 = \frac{\partial L}{\partial \dot{q}_1} = m\dot{q}_1 = m\dot{r}, \quad p_2 = \frac{\partial L}{\partial \dot{q}_2} = mr^2\dot{\varphi}.$$

Первое скалярное уравнение Лагранжа

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{r}} - \frac{\partial L}{\partial r} = 0$$

приводит к соотношению

$$m\ddot{r} = mr\varphi^2 - \frac{\partial U}{\partial r},$$

а второе — к соотношению $\partial L/\partial \dot{q}_2 = 0$, или $p_2 = \mathrm{const}$, — т.е. к закону сохранения обобщенного импульса. Здесь p_2 — первый интеграл системы дифференциальных уравнений Лагранжа.

Этот пример позволяет сделать следующее обобщение.

Определение. Координата q_i называется *циклической координатой*, если функция Лагранжа не зависит от q_i явно, т.е. если $\partial L/\partial q_i = 0$.

Теорема. *Обобщенный импульс, соответствующий циклической координате, сохраняется.*

Уравнения Лагранжа можно записать как систему $2N$ уравнений первого порядка:

$$p_i = \frac{\partial L}{\partial \dot{q}_i}, \quad \dot{p}_i = \frac{\partial L}{\partial q_i}, i = 1, \ldots, N.$$

Теорема. *Система уравнений Лагранжа эквивалентна системе $2N$ уравнений Гамильтона:*

$$\dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}}, \quad \dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}},$$

где $H(\mathbf{p}, \mathbf{q}, t) = \mathbf{p}\dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}}, t)$.

4.2.2. Законы сохранения и инвариантность гамильтониана. Если уравнения Лагранжа описывают механическую систему, и $L = T - U$, где $T = \sum a_{ij}\dot{q}_i\dot{q}_j/2$ — квадратичная форма относительно $\dot{q}_i$, а $U = U(q_1, \ldots, q_N)$, то H есть полная энергия системы: $H = T + U$.

Если $U = U(\mathbf{q})$, то справедлив закон сохранения энергии, который в общем виде формулируется следующим образом: если $\frac{\partial H}{\partial t} = 0$ (т.е. H явно не зависит от времени), то выполняется закон сохранения функции Гамильтона. Действительно,

$$\frac{dH}{dt} = \frac{\partial H}{\partial \mathbf{p}}\left(-\frac{\partial H}{\partial \mathbf{q}}\right) + \frac{\partial H}{\partial \mathbf{q}}\frac{\partial H}{\partial \mathbf{p}} + \frac{\partial H}{\partial t} = \frac{\partial H}{\partial t}.$$

Рассматривая центральное поле, мы заметили, что введением полярных координат задача свелась к одномерной. Оказывается, всякая симметрия задачи, математически выражающаяся как инвариантность гамильтониана по отношению к преобразованиям фазового пространства и позволяющая выбрать систему координат $\mathbf{q}$ так, чтобы H не зависела от некоторых координат, позволяет найти первые интегралы системы уравнений Гамильтона и свести задачу к задаче с меньшим числом координат. В силу эквивалентности Лагранжева и Гамильтонова описаний, можно дать следующее эквивалентное определение:

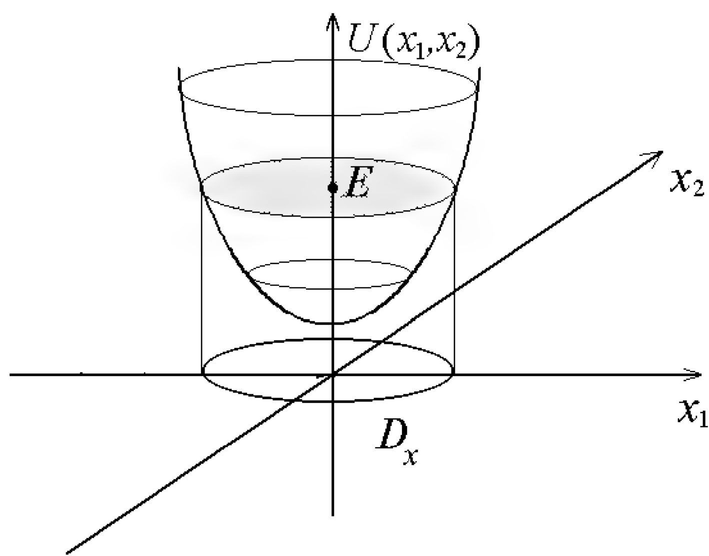


Рис. 4.2. Потенциал системы, для которой справедлива теорема Пуассона

Определение. Координата q_1 называется *циклической координатой*, если функция Гамильтона $H(p_1, \ldots, p_N; q_1, \ldots, q_N; t)$ не зависит от q_1 явно, т.е. если $\dfrac{\partial H}{\partial q_1} = 0$.

Следствие. *Пусть q_1 — циклическая координата, тогда p_1 — первый интеграл. При этом изменение остальных координат со временем описывается системой уравнений Гамильтона с независимыми координатами $\mathbf{q}' = (q_2, \ldots, q_N)$ и функцией Гамильтона $H(p_2, \ldots, p_N; q_2, \ldots, q_N; t; c)$, где $c = p_1$ — параметр.*

Действительно, обозначив $\mathbf{p}' = (p_2, \ldots, p_N)$, получим систему уравнений Гамильтона

$$\frac{d\mathbf{q}'}{dt} = \frac{\partial H}{\partial \mathbf{p}'}, \quad \frac{dq_1}{dt} = \frac{\partial H}{\partial p_1};$$

$$\frac{d\mathbf{p}'}{dt} = -\frac{\partial H}{\partial \mathbf{q}'}, \quad \frac{dp_1}{dt} = 0.$$

Отсюда $p_1 = \mathrm{const}$. После интегрирования системы $2N - 2$ уравнений для определения q_1 имеем уравнение

$$\dot{q}_1 = \frac{\partial}{\partial p_1} H(p_1, \mathbf{p}'(t); \mathbf{q}'(t), t),$$

которое легко интегрируется.

Следствие. *Всякая консервативная система с двумя степенями свободы, $N = 2$, имеющая циклическую координату, интегрируема, так как в этом случае система для q', p' — одномерна и интегрируется с помощью первого интеграла $H(p', q') = \mathrm{const}$.*

4.2.3. Особенности фазовых портретов гамильтоновых систем. В этом разделе формулируется ряд важнейших фактов теории гамильтоновых систем. Сначала определим основные понятия, характерные для динамики гамильтоновых систем.

Определение. $2N$-мерное векторное пространство $\mathcal{R}_{2N}$ с координатами $p_1, \ldots, p_N$; $q_1, \ldots, q_N$ называется *фазовым пространством.*

Определение. *Фазовым потоком* называется однопараметрическая группа преобразований фазового пространства $g^t(\mathbf{p}(0), \mathbf{q}(0)) = (\mathbf{p}(t), \mathbf{q}(t))$.

Теорема Лиувилля. *Для консервативной системы фазовый поток сохраняет фазовый объем, т.е. для любой области D фазового пространства имеем: объем $\{g^t D\}$ = объем $\{D\}$.*

Сформулируем еще одно важное утверждение, имеющее многочисленные применения.

Теорема Пуанкаре о возвращении. *Пусть g — сохраняющее объем непрерывное взаимно однозначное отображение,*

переводящее ограниченную область D евклидова пространства в себя: $gD = D$. Тогда в любой окрестности U любой точки области D найдется точка $x \in U$, которая возвращается в область U, т.е. $g^n x \in U$ при некотором $n > 0$.

Теорема Пуанкаре применима, например, к фазовому потоку g^t двумерной системы с растущим на бесконечности потенциалом $U(\mathbf{q}) = U(x_1, x_2)$, (рис. 4.2); в этом случае инвариантная (по отношению к фазовому потоку) область фазового пространства дается условием $D = \{\mathbf{p}, \mathbf{q} : \ T + U \leq E\}$.

Теорему Пуанкаре можно усилить, доказав, что почти всякая движущаяся точка многократно возвращается к своему исходному положению. Это — один из немногих общих выводов о характере движения, деталей которого (дающиеся точным решением дифференциальных уравнений движения), как правило, не удается получить.

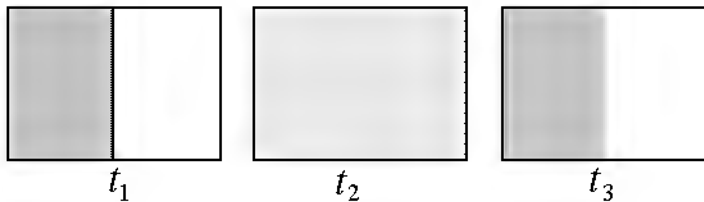


Рис. 4.3. Следствие теоремы Пуанкаре

Из теоремы Пуанкаре следует парадоксальный вывод: если открыть перегородку, отделяющую камеру с газом от вакуума, то молекулы газа через некоторое время вновь соберутся в первом отсеке, как это показано на рис. 4.3. “Разгадка” парадокса состоит в том, что для этого потребуется время, многократно превосходящее время жизни Вселенной.

Приведем несколько следствий из теоремы Пуанкаре.

1. Пусть D — окружность, g — поворот на угол α. Если $\alpha = 2\pi m/n$, где $m, n \in \mathbf{Z}$, т.е. целые числа, то g^n — тождественное преобразование. Если же α несоизмеримо, или, как говорят, арифметически независимо, с 2π, то теорема Пуанкаре дает: для любого $\delta > 0$ найдется такое n, что

$$| \, g^n x - x \, | < \delta.$$

Иными словами, это означает, что если $\alpha \neq 2\pi m/n$, $m, n \in \mathbf{Z}$, то множество точек вида $g^n x$, $x \in \mathbf{Z}$, всюду плотно на окружности (рис. 4.4).

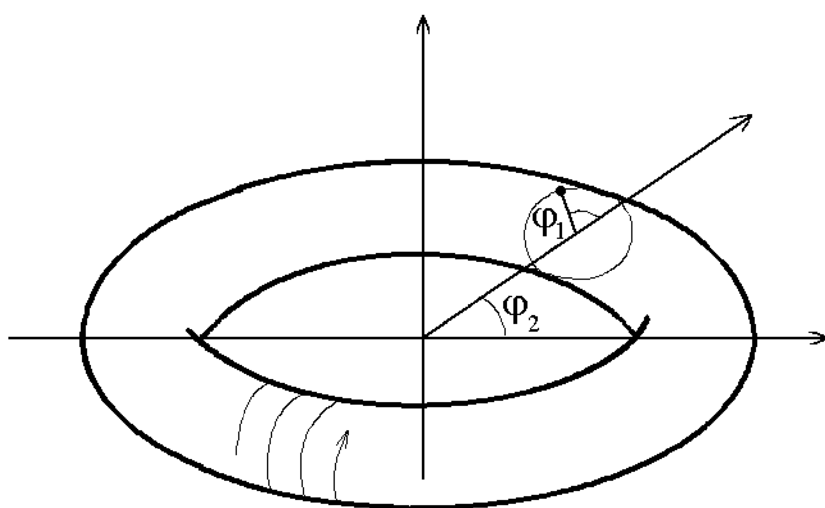


Рис. 4.4. Фазовая траектория всюду плотно заполняет тор

2. Пусть D — двумерный тор, φ_1 и φ_2 — угловые координаты на нем. Рассмотрим эволюционные уравнения

$$\dot{\varphi}_1 = \alpha_1, \quad \dot{\varphi}_2 = \alpha_2.$$

Ясно, что фазовый поток $g^t(\varphi_1, \varphi_2) = (\varphi_1 + \alpha_1 t, \varphi_2 + \alpha_2 t)$ сохраняет объем $d\varphi_1 \, d\varphi_2$. Из теоремы Пуанкаре следует, что если отношение

α_1/α_2 — иррационально, то "обмотка" тора $\{g^t(\varphi_1, \varphi_2), t > 0\}$ всюду плотна на торе.

Приведенные здесь факты являются основой для исследования вполне интегрируемых гамильтоновых систем и их возмущений.

4.3. Вполне интегрируемые системы

Рассмотрим общие качественные особенности поведения вполне интегрируемых гамильтоновых систем, т.е. таких систем, в которых известно ровно N первых интегралов, что позволяет, воспользовавшись результатами, изложенными в предыдущем разделе, получить решение гамильтоновой системы уравнений в виде квадратур.

4.3.1. Скобки Пуассона и первые интегралы. Напомним некоторые факты, известные из теоретической механики. Если гамильтониан не зависит явно от времени, $H(\mathbf{p}, \mathbf{q}, t) = H(\mathbf{p}, \mathbf{q})$, то система называется консервативной, и в ней выполняется закон сохранения гамильтониана (энергии), т.е. $H(\mathbf{p}, \mathbf{q}) = E = \mathrm{const}$, и можно считать известным один первый интеграл движения. Все остальные первые интегралы движения F системы с гамильтонианом $H(\mathbf{p}, \mathbf{q})$ удовлетворяют условиям

$$[F, H] = 0,$$

где $[\cdot, \cdot]$ — скобки Пуассона:

$$[A, B] = \sum_{i=1}^{N} \frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q_i} - \frac{\partial A}{\partial q_i}\frac{\partial B}{\partial p_i}.$$

Говорят, что функции A и B находятся в инволюции, если их скобки Пуассона равны нулю.

Одной из проблем динамики является отыскание первых интегралов, независимых от H (нетривиальных), т.е. таких, что $[F, H] = 0$ и $F \neq F(H)$. Найти их довольно трудно, но можно указать условия их существования.

Лиувиллем показано, что если в системе с N степенями свободы (и размерностью фазового пространства, равной $2N$), известны N независимых первых интегралов в инволюции, т.е. $[F_i, F_j] = 0$, $i \neq j$, $i, j = 1, \ldots, N$, то система интегрируется в квадратурах и ее эволюция подчиняется весьма общим закономерностям.

4.3.2. Условно периодическое движение.

Определение. Пусть T^N — N-мерный тор в фазовом пространстве, $\varphi = (\varphi_1, \ldots, \varphi_N)$ $[\mathrm{mod}\,(2\pi)]$ — угловые координаты на

торе T^N. *Условно периодическим движением* называется движение, описываемое уравнениями

$$\dot{\varphi}_i(t) = \omega_i, \quad i = 1, \ldots, N. \tag{4.6}$$

Числа ω_i, $i = 1, \ldots, N$, называются частотами условно периодического движения.

Уравнения (4.6) интегрируются:

$$\varphi_i(t) = \omega_i t + \varphi_i(0), \quad i = 1, \ldots, N,$$

и графики зависимости координат $\varphi(t)$ от времени изображаются прямыми линиями, как показано на рис. 4.5.

Частоты условно периодического движения называются независимыми (арифметически), если для целых $k_i \in \mathbf{Z}$, $i = 1, \ldots, N$, равенство

$$\sum_{i=1}^{N} k_i \omega_i = 0$$

возможно тогда и только тогда, когда $k_i = 0$, $i = 1, \ldots, N$.

Сформулируем теперь точное утверждение, известное как теорема Арнольда–Лиувилля.

Теорема. *Пусть дана гамильтонова система с гамильтонианом $H = H(\mathbf{p}, \mathbf{q})$ с N степенями свободы, совершающая финитное движение и имеющая N первых интегралов $F_i = F_i(\mathbf{p}, \mathbf{q})$, $i = 1, \ldots, N$, линейно независимых и находящихся в инволюции*:

$$[F_i, F_j] = 0, \quad i \neq j, \quad i, j = 1, \ldots, N.$$

Тогда:

— фазовые траектории системы лежат на N-мерном торе;

— движение системы условно периодическое и характеризуется N частотами $\omega_i = \omega_i(F_1, \ldots, F_N)$, $i = 1, \ldots, N$;

— угловые переменные $\varphi_1, \ldots, \varphi_N$, задающие координаты на торе, удовлетворяют уравнениям $\dot{\varphi}_i = \omega_i(F_1, \ldots, F_N)$, $i = 1, \ldots, N$, т.е. $\varphi_i(t) = \varphi_i(0) + \omega_i t$, $i = 1, \ldots, N$.

Таким образом, в условиях теоремы Арнольда–Лиувилля траектории системы всегда остаются на торе, поэтому в таких случаях говорят о существовании инвариантных (относительно фазового потока) торов в интегрируемом случае.

Пусть, например, $N = 2$ и известен первый интеграл F, не зависящий от H. Тогда система интегрируема в квадратурах, траектория его лежит на инвариантном торе и ее движение условно периодично. Фазовые траектории, отвечающие фиксированному уровню энергии (гамильтониана), принадлежат трехмерному пространству, и торы, отвечающие различным значениям интеграла F, являются вложенными один в другой и делят все пространство, (рис. 4.6). При

$N > 2$ торы не имеют столь простой структуры, взаимно проникают друг в друга и не делят фазовое пространство.

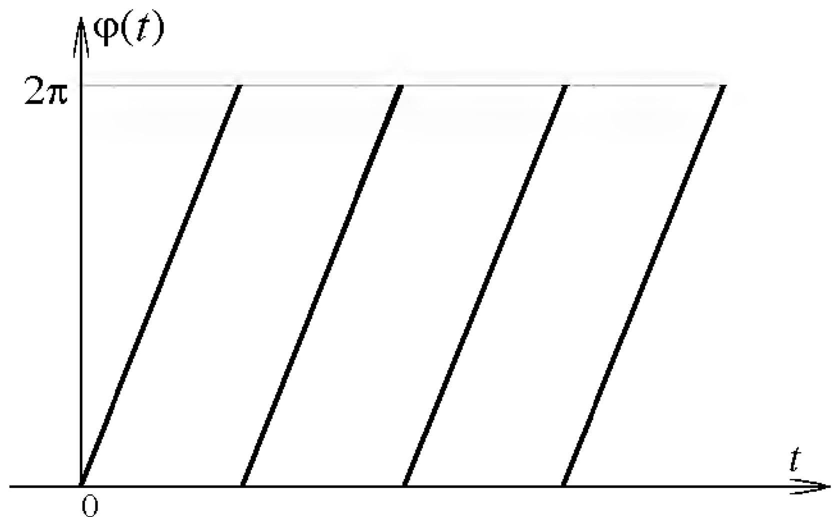
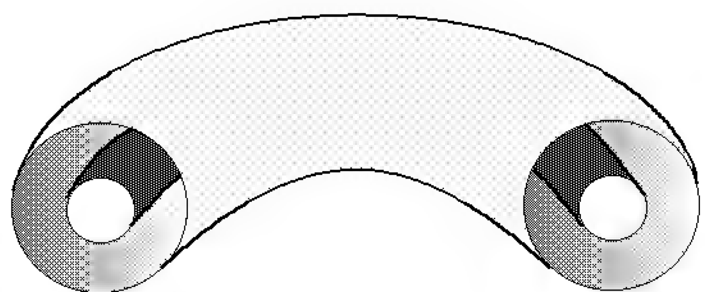


Рис. 4.5. Зависимость координаты условно периодического движения от времени

Итак, из теоремы Арнольда–Лиувилля следует, что множество $M_f = \{x \in \mathcal{R}_{2N} : F_i(x) = f_i, i = 1, \ldots, N\}$ точек фазового пространства $\mathcal{R}_{2N}$ есть N-мерный тор, инвариантный относительно фазового потока. На нем можно выбрать фазовые координаты так, что фазовый поток с гамильтонианом $H = F_1$ принимает на M_f простой вид:

$$\frac{\partial \varphi}{\partial t} = \omega(t).$$

Интерес представляет выбор оставшихся N координат в окрестности N-мерного многообразия M_f в $\mathcal{R}_{2N}$ так, чтобы получить набор канонических координат, т.е. таких, для которых эволюционные уравнения оставались бы гамильтоновыми. Сами интегралы F_i в общем случае не могут служить каноническими координатами, однако можно ввести некоторые функции от них, $I_i = I_i(F_1, \ldots, F_N)$, $i = 1, \ldots, N$, такие, что $(\mathbf{I}, \varphi)$ уже будут каноническими. Это так называемые переменные действия, процедура их введения известна из теоретической механики.


Рис. 4.6. Двумерные торы "делят" трех - мерное пространство

В переменных "действие–угол" система уравнений Гамильтона принимает весьма простой вид:

$$\dot{I}_i = 0, \quad \dot{\varphi}_i = \omega_i, \quad i = 1, \ldots, N.$$

Движение фазовой точки по инвариантному тору $\mathbf{I} = \mathrm{const}$ является условно периодическим с частотами

$$\omega_i = \frac{\partial H}{\partial I_i}, \quad i = 1, \ldots, N.$$

4.3.3. Резонансные и нерезонансные торы в фазовом пространстве. Вернемся к примеру системы с двумя степенями свободы. Если частоты ω_1 и ω_2 несоизмеримы, то, как следует из теоремы Пуанкаре, фазовая траектория всюду плотно обвивает инвариантный

тор. Если же $\omega_1/\omega_2 = m_1/m_2$, где m_1, $m_2 \in \mathbf{Z}$, то через конечное число оборотов, равное m_2, траектория замкнется. Этот случай называется резонансным, в отличие от нерезонансного при иррациональном отношении частот условно периодического движения. Резонансный случай производит впечатление вырождения, так как мера рациональных чисел на прямой равна нулю (рациональных чисел "меньше", чем иррациональных). Несмотря на это, сколь угодно малым изменением частот мы можем попасть с резонансного случая на нерезонансный и наоборот, это происходит потому, что множество рациональных чисел всюду плотно на прямой.

В общем случае, когда $N > 2$, резонансное вырождение связывают с существованием ненулевого N-мерного вектора с целочисленными координатами $\mathbf{m} = (m_1, \ldots, m_N)$, такого, что

$$\sum_{i=1}^{N} m_i \omega_i = 0. \tag{4.7}$$

Если соотношение (4.7) выполнено для $r \le N$ различных (арифметически независимых) векторов $\mathbf{m}$, то говорят об r-кратном вырождении. Фазовая кривая всюду плотно заполняет тор такого числа измерений, сколько среди частот $\omega_1, \ldots, \omega_N$ арифметически независимых (частоты арифметически независимы, если равенство (4.7) возможно только при $\mathbf{m} = 0$).

Нерезонансные торы в $2N$-мерном фазовом пространстве образуют множество полной меры, в то время как мера Лебега объединения всех резонансных торов равна нулю. В то же время резонансные торы переплетаются с нерезонансными и образуют в фазовом пространстве всюду плотное множество. Более того, всюду плотно любое множество всех резонансных торов с фиксированным числом резонансных частот (от 1 до $N - 1$).

4.3.4. О теории Колмогорова–Арнольда–Мозера. В такой ситуации чрезвычайно важна проблема устойчивости резонансных и нерезонансных торов по отношению к консервативному возмущению гамильтониана. Этот вопрос возникает при прогнозе движения заряженных частиц в электрическом и магнитном полях в связи с решением задач об удержании плазмы в различного рода ловушках. Эволюция планет Солнечной системы также может считаться условно периодическим движением в центрально симметричном поле только если пренебречь притяжением планет между собой. Ответ на этот вопрос дает теория Колмогорова–Арнольда–Мозера (КАМ-теория).

Рассмотрим систему, описываемую "возмущенным" гамильтонианом

$$H = H_0(I_1, \ldots, I_N) + \varepsilon V(I_1, \ldots, I_N; \varphi_1, \ldots, \varphi_N), \quad \varepsilon \ll 1.$$

Невозмущенный гамильтониан H_0 описывает условно периодическое движение на N-мерном торе, при $\varepsilon = 0$ торы инвариантны. Что можно сказать о движении системы при "включении" возмущения (не ограничиваясь временными рамками)?

Введем понятие функциональной независимости частот

Определение. Частоты $\omega_i(\mathbf{I})$, $i = 1, \ldots, N$, называются *функционально независимыми*, если

$$\det\left(\frac{\partial\omega_i}{\partial I_j}\right) = \det\left(\frac{\partial^2 H_0}{\partial I_i \partial I_j}\right) \neq 0. \tag{4.8}$$

Это понятие удобно проиллюстрировать на примере системы с двумя степенями свободы (четырехмерное фазовое пространство).

Рассмотрим трехмерное многообразие фазового пространства постоянной энергии. Оно расслаивается на двумерные торы, на каждом из которых оба первых интеграла постоянны. Построим площадку, трансверсально (т.е. под ненулевым углом для каждой фазовой траектории на торе) пересекающую торы. Фазовая кривая, начинаясь на такой площадке, вновь ее пересекает, обернувшись вокруг тора — в результате имеется новая точка на окружности, по которой площадка пересекает тор. Тем самым возникает отображение площадки на себя (отображение Пуанкаре). Указанное отображение оставляет на месте концентрические меридианные окружности, по которым площадка пересекает торы, при этом окружность поворачивается на некоторый угол, а именно, на такую долю полного оборота, какую частота вдоль меридиана тора составляет от частоты вдоль экватора. Если система невырождена (т.е. выполнено соотношение (4.8)), то угол поворота инвариантных окружностей на плоскости сечения меняется от окружности к окружности.

В зависимости от соизмеримости и несоизмеримости частот образ любой точки окружности либо возвращается в нее через некоторое целое число оборотов, либо всюду плотно покрывает окружность. В последнем случае нет ни одной неподвижной точки отображения окружности в себя, в то время как в случае соизмеримости частот (резонансный случай) имеем полную окружность неподвижных точек (весьма противоестественное поведение для отображения сколько-нибудь общего вида). При сколь угодно малом возмущении (общего вида) это свойство отображения должно пропасть — окружность неподвижных точек должна рассыпаться (оставив, возможно, лишь конечное их число).

Итак, можно ожидать, что при малом возмущении интегрируемой системы происходит качественное изменение картины движения (бифуркация), при котором инвариантные торы, заполненные замкнутыми траекториями, должны рассыпаться, оставив лишь конечное число

замкнутых кривых, а остальные фазовые кривые будут вести себя сложнее.

Что же произойдет с нерезонансными торами? Колмогорову и Арнольду принадлежит теорема о сохранении нерезонансных инвариантных торов при независимых функционально частотах периодического движения (так называемый невырожденный случай). Оказывается, что при включении консервативного малого возмущения в невырожденном случае в фазовом пространстве большинство инвариантных торов не исчезает, а лишь немного деформируется так, что в фазовом пространстве возмущенной системы тоже имеются инвариантные торы, заполненные всюду плотно фазовыми кривыми, обматывающими их условно периодически с числом частот, равным числу степеней свободы. Указанные инвариантные торы образуют большинство в том смысле, что мера их дополнения мала при малых возмущениях.

Что же происходит в тех случаях, когда начальные условия попадают в щель между инвариантными торами, которая возникла на месте распавшегося резонансного тора?

В случае $N = 2$ уровни энергии $H = E$ выделяют трехмерные пространства, в которых существуют инвариантные торы, делящие все пространство. Поэтому траектория, начинающаяся в щели между двумя инвариантными торами, оказывается запертой между ними, и соответственно переменные действия всегда оказываются вблизи своих начальных значений.

В случае $N > 2$ условно периодические инвариантные торы не делят все пространство, поэтому фазовые траектории по щели могут "путешествовать" по всему фазовому пространству как угодно далеко от начального состояния по переменным действия. Скорость ухода этих переменных была оценена и имеет вид

$$e^{-1/\varepsilon^d}, \quad d \in (0,1),$$

т.е. эта скорость экспоненциально мала и не может улавливаться методами теории возмущений.

4.4. Инвариантные торы в негамильтоновых системах

Заметим, что инвариантные торы могут возникать не только в консервативных гамильтоновых системах.

П р и м е р (генератор Ван дер Поля с возбуждением). Пусть

$$\ddot{x} - a(1 - bx^2)\dot{x} + x = B\sin(pt + \varphi_0),$$

где амплитуда возбуждения B достаточно мала, а частота возбуждения p несоизмерима с частотой автоколебания автономного осциллятора. Периодическая модуляция предельного цикла автономной системы приводит к тому, что фазовая траектория с частотой p вращается вокруг предельного цикла и лежит на двумерном торе. Аналогично случаю автономной системы, это множество является устойчивым предельным множеством. Минимальная размерность фазового пространства, в которое можно вписать двумерный тор, равна трем, поэтому системы с гармоническим возбуждением называют системами с числом степеней свободы, равным 3/2.

Глава 5

Хаос в динамических системах

5.1. Что такое хаос

5.1.1. Хаос: мифология и математика. Хаос в древних философских теориях означал бесконечное пространство, существовавшее до начала мира. В греческой мифологии это беспорядочная субстанция, из которой возник порядок — Вселенная, из нее вышли Боги, люди, Земля, небесные светила. На протяжении нескольких тысячелетий это понятие было достоянием философии и мифологии, науке же предназначалось описание "упорядоченного мира", т.е. Космоса в понимании античных философов.

В современном мире с хаосом связывается неповторяющаяся, нерегулярная, беспорядочная последовательность состояний. Буквально несколько десятилетий назад считалось, что такие процессы крайне редки, природа же развивается непрерывно, без резких скачков. Действительно, вся классическая физика — механика Ньютона и Галилея, электродинамика Максвелла, статистическая физика, отчасти и современная — квантовая теория, например, оперирует с понятием функции и отображения, геометрическим образом которого является кривая или поверхность. Галилею принадлежит фраза: "Вся наука записана в великой книге — я имею в виду Вселенную, — которая всегда открыта для нас, но которую нельзя понять, не научившись понимать язык, на котором она написана. А написана она на языке математики, и ее буквами являются треугольники, окружности и другие геометрические фигуры, без которых человеку невозможно разобрать ни одного ее слова; без них он подобен блуждающему во тьме". Во времена Галилея под функцией понималось лишь то, что в современной математике называют непрерывной функцией, ее график можно нарисовать, не отрывая пера от бумаги. Такой подход к описанию природы заранее исключал возможность рассмотрения

полного беспорядка — хаоса.

Однако с развитием понятия функции усложнялись и геометрические образы, которыми оперировали физики для описания природы. Достаточно сложные математические объекты, такие, например, как функция, имеющая разрыв в каждой точке (функция Дирихле), непрерывная линия, плотно заполняющая весь квадрат, или множество точек плоскости, не имеющее площади, стали рассматриваться примерно сто лет назад. Геометрические образы этих математических объектов довольно трудно представить и невозможно нарисовать. Эти примеры абстрактных объектов могут показаться пустой игрой ума, однако есть и примеры природных образований, явлений и процессов, которые требуют для описания привлечения математических объектов со столь экзотическими свойствами; они получили название фракталов. Эти объекты и лежат в основе современной теории хаотических процессов.

Почему хаос казался экзотикой несколько лет назад? Потому что эволюцию систем со времен Лапласа принято описывать, задавая начальное состояние и скорости его изменения, для чего и была создана прекрасно работающая на практике теория дифференциального исчисления. С математической точки зрения, поведение системы в любой момент времени полностью определено, если выполняются условия существования и единственности решения соответствующего дифференциального уравнения. Долгое время считалось, что в такой определенной (детерминированной) системе не может возникать хаоса, ведь решение этого уравнения — “гладкая”, т.е. непрерывная и дифференцируемая функция, и лишь на границе XIX и XX веков Анри Пуанкаре обнаружил, что в некоторой гамильтоновой механической системе могут появляться хаотические движения. Эти примеры были восприняты современниками как парадокс.

Однако сейчас стало совершенно ясно, что если речь идет о достаточно сложной нелинейной системе, то ее хаотическое состояние скорее правило, чем исключение, и состояние хаоса является неотъемлемой частью таких реальных систем. К настоящему времени открыто множество примеров динамических систем, в которых возникают состояния нерегулярного, хаотического движения. Их прекрасной иллюстрацией служат забавные механические игрушки, появившиеся сейчас в продаже — маятники на карданных подвесах, вычурные движения которых приковывают к себе взгляд и завораживают, подобно текущей воде или огню костра. Подчеркнем, что такое поведение не является ни следствием случайного возмущающего воздействия — такие воздействия не включены в модель системы, приходящей к хаосу, ни следствием бесконечного числа степеней свободы — хаос возникает уже в системах, описываемых тремя координатами (или в терминах раздела 4.4, обладающими

полутора степенями свободы), ни следствием неопределенности (классической или квантовой) в начальных данных. Причина появления хаотических режимов лежит в нелинейной природе динамической системы и в свойстве их неустойчивости, проявляющемся в необычайно быстром экспоненциальном разбегании первоначально близких траекторий: при достаточно большом удалении состояния системы от начального включаются нелинейные механизмы, возвращающие траекторию в окрестность начальной точки (см. теорему Пуанкаре о возвращении), неустойчивость ее вновь отбрасывает, и за счет этого происходит "беспорядочное" запутывание траектории. Заметим, что в линейных моделях, характерных для науки XVII–XIX веков, да даже и начала нашего столетия, не возникает хаотических режимов, они являются исключительным свойством нелинейных систем.

Интересно, что теоретически хаотическая траектория полностью воспроизводится, если создать точно такие же начальные условия, однако свойство неустойчивости динамической системы тесно связано с непредсказуемостью их поведения на большой период времени. На практике это означает, что невозможно предсказать поведение хаотической системы на большой период времени, так как начальные условия и вычисления можно проводить лишь с определенной точностью; по сути дела, это свойство хаотических систем — необычайная чувствительность к малым воздействиям — означает конец эпохи лапласовского детерминизма. Одно из далеко идущих следствий этого свойства иллюстрируется примером так называемой "бабочки Лоренца": взмах крыльев бабочки может повлиять на климат Земли в глобальном масштабе, так как атмосфера является сложной нелинейной системой с неустойчивыми режимами.

Такая ситуация приводит и к другим интересным выводам, чрезвычайно важным как с теоретической, так и с практической точки зрения. Например, оказывается, что сложная нелинейная система в процессе своего развития обязательно проходит через этапы хаоса. В физике такими этапами являются так называемые фазовые переходы. Человек — тоже сложная нелинейная система, и всем нам знакомы кризисы и депрессии, своеобразные периоды хаоса. Развитие общества проходит через этапы социальных, технических, экономических и других революций, также сопровождающихся хаосом. Кризисы и революции — обязательные этапы развития, и если вы хотите оставаться застывшими, неподвижными, вы обязаны проходить через хаос. Изучение математических моделей хаоса учит нас относиться к этому не как к катастрофе, а как к естественному природному явлению.

Возвращаясь к мифологии, можно вспомнить, что хаос призван уничтожить, поглотить, разрушить старое, отжившее, и дать дорогу новому, не существовавшему прежде, причем зародыши, точнее,

никак еще не проявленные потенциалы этого нового содержатся в самом хаосе, являясь его природой. Находясь в хаосе, имеет смысл не теряться, а постараться уловить основную тенденцию хаотического процесса, и постараться ему следовать — в этом неоценимую роль может сыграть математическая модель. Свойства хаоса таковы, что если изменение состояния системы совпадает по направлению с потенциальным путем развития, предназначенным законами динамики, то он приводит к чрезвычайно значимым, заметным результатам, сколь бы мало это движение ни было в начале. Математически это обусловлено существованием неустойчивых направлений возмущения нелинейной системы. Казалось бы, мы вновь пришли к предопределенности? Не совсем так, ведь неустойчивых направлений много, и система вольна выбрать любое из них.

Взрывной первоначальный рост новой формы не может продолжаться вечно. Рано или поздно включатся стабилизирующие силы, связанные с нелинейностью. Они гармонизируют систему, стабилизируют ее и дают возможность спокойного существования в течение достаточно длительного периода времени. Это — период накопления опыта, осознания роли нового рождения, период "выполнения миссии". Это время устойчивого развития характеризуется тем, что теперь практически невозможно переключиться на иной режим, существенно отличающийся от данного. В греческой мифологии такие этапы символически связывались с двумя божествами, Дионисом и Аполлоном, первый из них — бог творческого вдохновения, экстаза, растворяющийся во множестве рожденных форм, а второй — бог гармонии, успокаивающий, приводящий в порядок, отбрасывающий все лишнее.

Символическое видение мира, закрепленное и дошедшее до нас в мифологических образах, удивительным образом смыкается с современной наукой, использующей в своем языке иные образы и символы — математические.

5.1.2. Хаотические колебания. Перейдем к формальному описанию хаотической динамики, и для этого начнем с простых примеров.

Маятник с возбуждающей силой. Рассмотрим материальную точку, движение которой вдоль прямой описывается эволюционным уравнением

$$\ddot{\varphi} + \gamma\dot{\varphi} + g\sin\varphi = F\cos(\omega t).$$

Как упоминалось в разделе 4.4, эта система с 3/2 степенями свободы; этот термин можно оправдать также тем, что можно ввести три обобщенные координаты, описывающие ее эволюцию: $x = \varphi$, $y = \dot{\varphi}$,

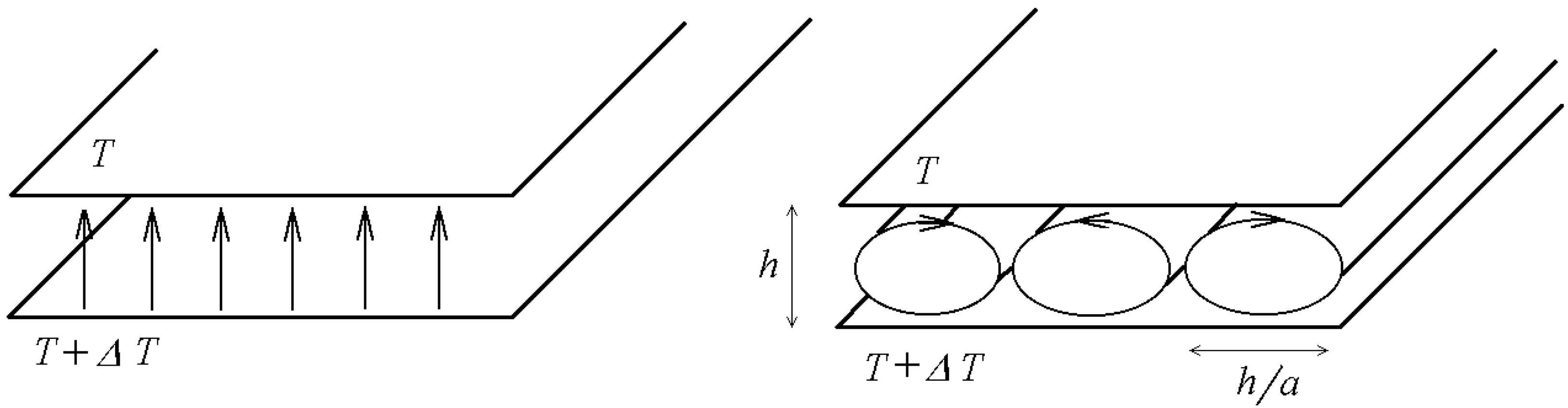


Рис. 5.1. Возникновение ячеек Бенара в двумерной среде

$z = \omega t$. Тогда эволюционное уравнение примет вид

$$\dot{x} = y,$$
$$\dot{y} = -\gamma y - g\sin x + F\cos z,$$
$$\dot{z} = \omega,$$

и при любом начальном условии при достаточно больших значениях амплитуды при несоизмеримых собственной частоте и частоте F возбуждающей силы зависимость угла φ от времени выглядит хаотической.

Ячейки Бенара. Рассмотрим слой жидкости (или газа — если речь идет об атмосферных процессах), находящейся в поле тяжести и подогреваемый снизу.

Нагретая жидкость стремится подняться от дна к поверхности, а холодная — опуститься вниз. Но этому движению противостоит сила вязкого трения, и если разность температур ΔT нижнего и верхнего слоев мала, так, что преобладают механизмы трения, то тепло передается только посредством теплопроводности. Если же эта разность превысит некоторый критический порог, то состояние покоя становится неустойчивым и возникают стационарные конвективные потоки (ячейки Бенара), рис. 5.1. Это явление — образование устойчивых потоков в среде с градиентом температуры, направленным против поля тяжести — знакомо каждому, кто наблюдал за разогретым маслом на сковороде, и обращал внимание на странный устойчивый рисунок, в виде шестиугольных сот, образованный кипящим в ячейках Бенара маслом. Этот рисунок постоянно возникает, несмотря на ровную и чистую (а значит — однородную и изотропную) поверхность сковороды. С дальнейшим ростом градиента тепла происходит разрушение этих валов и эволюция системы становится хаотической. Параметр R, описывающий качественное поведение динамической системы, называется параметром Рэлея и зависит от баланса механизмов вязкости и теплопроводности; он пропорционален градиенту температуры.

5.1.3. Аттрактор Лоренца. Вывод уравнений. Упрощенный анализ возникновения и разрушения конвективных вихрей в среде

привел Лоренца к ставшей уже классической модели динамической системы, описываемой тремя обыкновенными дифференциальными уравнениями и получившей название аттрактора Лоренца.

Жидкость или газ, образующий вихри в ячейках Бенара, является динамической системой с распределенными параметрами и описывается дифференциальными уравнениями в частных производных. Лоренцу удалось провести удачную параметризацию этой системы, сведя ее к системе обыкновенных дифференциальных уравнений, описывающих параметры вихрей.

Поясним, как получены уравнения аттрактора Лоренца. В качестве характеристик исходной системы рассмотрим поле скоростей $\mathbf{V}(\mathbf{x}, t)$ и поле температур $T(\mathbf{x}, t)$. Запишем полную систему уравнений гидродинамики, которая состоит из следующих уравнений.

1. Уравнение Навье–Стокса:

$$\rho \frac{d\mathbf{V}}{dt} = \mathbf{F} - \nabla p + \mu \nabla^2 \mathbf{V}. \tag{5.1}$$

2. Уравнение теплопроводности:

$$\frac{\partial T}{\partial t} = k \nabla^2 T. \tag{5.2}$$

3. Уравнение неразрывности:

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \mathbf{V}) = 0. \tag{5.3}$$

Дополним систему уравнений граничными условиями

$$T(x, y, z, t)\,|_{z=0} = T_0 + \Delta T, \quad T(x, y, z, t)\,|_{z=h} = T_0. \tag{5.4}$$

Здесь ρ — плотность, μ — вязкость, p — давление среды, k — температуропроводность, $\mathbf{F} = \rho g \mathbf{e}_z$ — сила тяжести. Нелинейность системы заключена в полной производной в (5.1), так как

$$\frac{d\mathbf{V}}{dt} = \frac{\partial \mathbf{V}}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial \mathbf{V}}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial \mathbf{V}}{\partial z}\frac{\partial z}{\partial t} + \frac{\partial \mathbf{V}}{\partial t} = (\mathbf{V}, \nabla)\mathbf{V} + \frac{\partial \mathbf{V}}{\partial t}.$$

Упростим модель, предполагая, что поля температуры и скорости не зависят от y, и коэффициенты уравнений (5.1)–(5.3) являются константами, кроме плотности, для которой будем считать выполненным соотношение

$$\rho = \overline{\rho}(1 - \alpha \Delta T), \quad \overline{\rho} = \text{const}.$$

Запишем $V_x = u$, $V_y = w$, тогда уравнение (5.3) примет вид

$$\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} = 0. \tag{5.5}$$

Для того, чтобы это уравнение удовлетворялось автоматически, введем функцию $\psi(x, z, t)$, для которой выполнено

$$u = -\frac{\partial \psi}{\partial z}, \quad w = \frac{\partial \psi}{\partial x}, \tag{5.6}$$

тогда подстановка соотношений (5.6) в (5.5) даст тождество.

Для преобразования уравнения (5.2) введем функцию $\vartheta(x, z, t)$ отличия профиля температуры от линейного:

$$T(x, z, t) = T_0 + \Delta T - \frac{\Delta T}{h} z + \vartheta(x, z, t).$$

После этого уравнения (5.1)–(5.3) перейдут в уравнения

$$\frac{\partial}{\partial t} \nabla^2 \psi = -\frac{\partial(\psi, \nabla^2 \psi)}{\partial(x, z)} + \nu \nabla^4 \psi + g\alpha \frac{\partial \vartheta}{\partial x},$$

$$\frac{\partial \vartheta}{\partial t} = -\frac{\partial(\psi, \vartheta)}{\partial(x, z)} + \frac{\Delta T}{h} \frac{\partial \psi}{\partial x} + k \nabla^2 \vartheta,$$

где

$$\frac{\partial(a, b)}{\partial(x, z)} = \frac{\partial a}{\partial x} \frac{\partial b}{\partial z} - \frac{\partial a}{\partial z} \frac{\partial b}{\partial x}, \quad \nabla^4 = \frac{\partial^4}{\partial x^4} + \frac{\partial^4}{\partial z^4},$$

$\nu = \mu/\overline{\rho}$ — кинематическая вязкость. Эти уравнения получены применением оператора ротора к уравнению Навье–Стокса. Дополним граничные условия (5.4) условиями

$$\vartheta(0,0,t) = \vartheta(0,h,t) = \psi(0,0,t) = \psi(0,h,t) = \\ = \nabla^2 \psi(0,0,t) = \nabla^2 \psi(0,h,t) = 0,$$

и рассмотрим основные гармоники для функций ψ и ϑ, записав их представления в виде

$$\psi = \sqrt{2} k \frac{1 + a^2}{a} X(t) \sin\left(\frac{\pi a}{h} x\right) \sin\left(\frac{\pi}{h} z\right),$$

$$\vartheta = \frac{R_a \Delta T}{\pi R} \left(\sqrt{2} Y(t) \cos\left(\frac{\pi a}{h} x\right) \sin\left(\frac{\pi}{h} z\right) - Z(t) \sin\left(\frac{2\pi}{h} z\right) \right),$$

где амплитуды гармоник X, Y, Z зависят только от времени, число Рэлея $R = (g\alpha h^3 \Delta T)/(\nu k)$, $R_a = \pi^4 a^{-2} (1 + a^2)^3$ — критическое значение числа Рэлея, при котором начинается конвекция, a — отношение геометрических размеров вихря и глубины слоя (см. рис. 5.1).

Отсюда получается система обыкновенных дифференциальных уравнений для амплитуд:

$$\begin{aligned} \dot{X} &= -\sigma X + \sigma Y, \\ \dot{Y} &= -XZ + rX - Y, \\ \dot{Z} &= XY - bZ, \end{aligned} \tag{5.7}$$

где точкой обозначено дифференцирование по безразмерному времени $\tau = (\pi^2/h^2)(1+a^2)kt$, $\sigma = \nu/k$ — число Прандтля, $b = 4(1+a^2)^{-1}$, $r = R/R_a \sim \Delta T$ — управляющий параметр.

5.1.4. Анализ системы уравнений Лоренца. Проанализируем поведение системы (5.7). Для этого заметим, что нулями левой части (5.7) являются точки $\mathbf{X}_1 = (X_1, Y_1, Z_1) = (0,0,0)$, $\mathbf{X}_{2,3} = (\pm\sqrt{b(r-1)}, \pm\sqrt{b(r-1)}, r-1)$.

Первая точка соответствует состоянию теплопроводности без конвекции. Действительно, после линеаризации правой части (5.7) получим матрицу

$$\begin{pmatrix} -\sigma & \sigma & 0 \\ r & -1 & 0 \\ 0 & 0 & -b \end{pmatrix}$$

с собственными значениями

$$\lambda_{1,2} = -\frac{\sigma+1}{2} \pm \frac{1}{2}\sqrt{(\sigma+1)^2 + 4(r-1)\sigma}, \quad \lambda_3 = -b.$$

Эти собственные значения отрицательны при $0 < r < 1$, что соответствует устойчивости этого состояния равновесия при указанных значениях управляющего параметра r.

При $r = 1$ начинается конвекция Бенара, так как именно при этом значении управляющего параметра становятся устойчивыми точки $X_{2,3}$. Действительно, матрица линеаризации правой части (5.7) в этих точках имеет вид

$$\begin{pmatrix} -\sigma & \sigma & 0 \\ 1 & -1 & c \\ c & c & -b \end{pmatrix},$$

где $c = \pm\sqrt{b(r-1)}$, и ее собственные значения являются корнями полинома $P(\lambda) = \lambda^3 + (\sigma + b + 1)\lambda^2 + b(\sigma + r)\lambda + 2b\sigma(r-1)$, графики этого полинома и расположение корней при различных значениях управляющего параметра r приведены на рис. 5.2.

Если $r > r_1$, то два собственных числа становятся комплексными, т.е. появляется два предельных цикла, устойчивых до тех пор, пока действительные части этих собственных чисел неположительны.

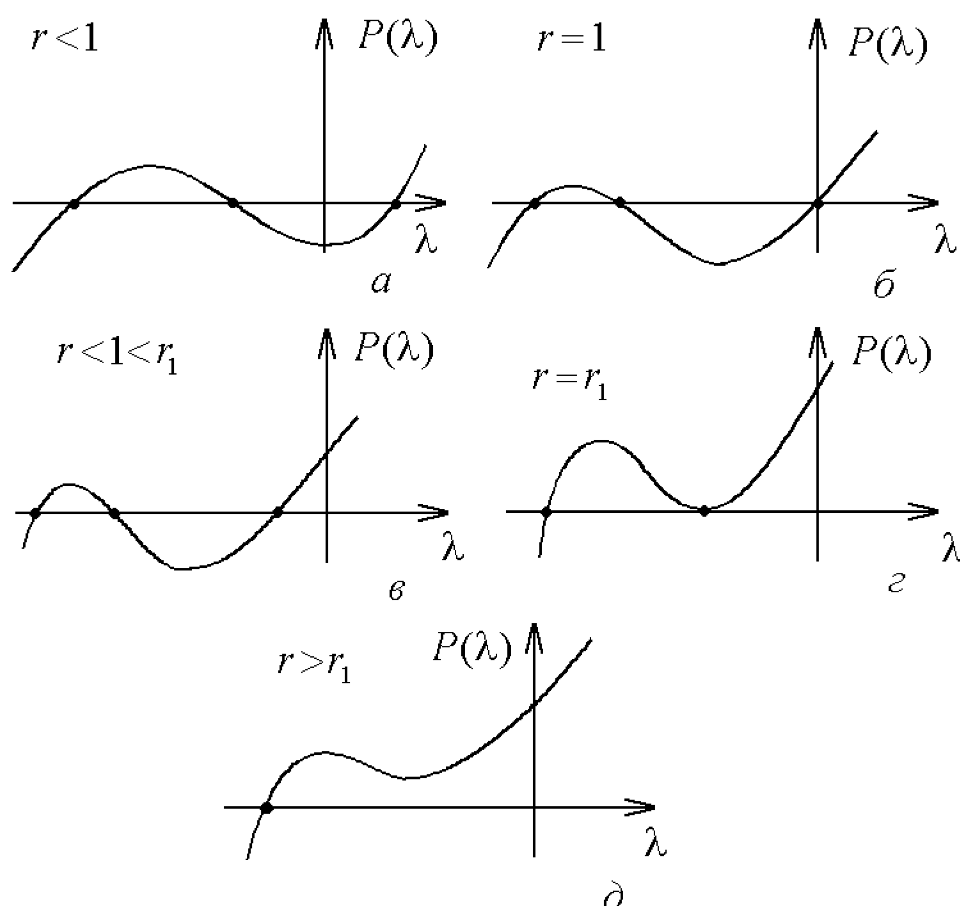


Рис. 5.2. Расположение корней характеристического уравнения при различных значениях управляющего параметра

При $r = r_{kp}$ действительные части обращаются в нуль, и при дальнейшем росте r наступает хаос (предельные циклы — неустойчивы).

Модель Лоренца адекватно описывает возникновение конвенции лишь в небольшой области изменения управляющего параметра вблизи перехода от теплопроводности к конвекции, так как для реальной гидродинамической системы при приближении к хаотическому режиму характерен рост амплитуды более высоких гармоник, не учитываемых в модели (5.7). Выполаживание спектра мощности нелинейных колебаний, когда в спектре вместо отдельных частот возникает целые полосы, свидетельствует о хаотическом поведении системы.

Нелинейная система (5.7) имеет и самостоятельный интерес, так как являет собой пример странного аттрактора — одного из необычных объектов теории нелинейных систем.

5.1.5. Реакция Белоусова–Жаботинского. Еще один характерный пример системы, демонстрирующей хаотическое поведение — это система химической кинетики, к которой относится реакция Белоусова–Жаботинского, также описывающаяся обыкновенными дифференциальными уравнениями относительно концентраций реагентов. При определенной концентрации ионов Ce^{4+} эта реакция представляет собой красочное зрелище — она проявляется в виде чередования движущихся цветных полос, визуализирующих волны концентрации реагентов, или в виде периодического изменения цвета раствора в колбе. При превышении некоторого порога концентрации Ce^{4+} все краски смешиваются — наступает хаос. Характерным сигналом хаотического режима реакции является уменьшение времени корреляции случайного поля концентраций.

5.1.6. Хаос и сечение Пуанкаре. Как уже неоднократно было замечено, удобным инструментом анализа динамических систем является отображение Пуанкаре. Если точки пересечения фазовой траектории с плоскостью Пуанкаре плотно и беспорядочно заполняют всю плоскость сечения, то это свидетельствует о чрезвычайной запу-

танности фазовой траектории — о хаотичном поведении системы.

5.1.7. Характерные признаки хаоса. Хаос — одно из важных и очень общих понятий, которому трудно дать математически строгое определение, однако, можно попытаться описать его характерные черты. Исходя из анализа приведенных в этом разделе примеров, можно выделить четыре характерных признака хаоса:

1). беспорядочное, нерегулярное поведение фазовой траектории;

2). широкополосный спектр мощности процесса $\mathbf{x}(t)$, где $\mathbf{x}$ — вектор координат динамической системы;

3). малый радиус корреляции процесса $\mathbf{x}(t)$;

4). плоскость сечения Пуанкаре плотно и нерегулярно заполнена точками пересечения с фазовой траекторией.

Несмотря на то, что эти критерии, скорее, являются описательными, чем математическими, тем не менее они дают представление о динамическом хаосе в нелинейных системах.

5.2. Дискретные отображения

Реальные нелинейные системы, как правило, весьма сложны для того чтобы разобраться с причинами и механизмами хаоса, поэтому мы начнем знакомство с ним с простых модельных систем, с одной стороны, поддающихся анализу методами математики, а с другой — несущих в себе все элементы реальных динамических систем.

В предыдущем разделе беспорядочность расположения точек "протыкания" фазовой траектории некоторой поверхности в фазовом пространстве была названа одним из признаков хаоса, следовательно, хаос динамической системы может проявится в хаотической последовательности преобразования Пуанкаре. Для того, чтобы лучше понять природу динамического хаоса, мы рассмотрим простые одномерные отображения Пуанкаре вида

$$x_{n+1} = f(x_n), \quad n = 1, 2, \ldots,$$

и с их помощью исследуем механизмы возникновения хаоса в нелинейной системе.

5.2.1. Сдвиг Бернулли. Рассмотрим одномерное дискретное отображение, называемое *сдвигом Бернулли*:

$$x_{n+1} = \sigma(x_n) \equiv 2x_n \pmod{1}, \quad x \in [0, 1], \tag{5.8}$$

с некоторым начальным условием $x_0 \in (0, 1)$. График отображения (5.8) приведен на рис. 5.3.

Запишем двоичное представление последовательности $x_0, x_1, \ldots$, полученное при начальном значении $x_0 = \sum_{k=1}^{\infty} a_k 2^{-k}$, где a_k принимают значения либо нуль, либо единица. Действие преобразования Бернулли на двоичную дробь приводит к удалению первого знака после запятой и сдвигу оставшейся последовательности влево (сдвиг Бернулли). Рассмотрим свойства преобразования (5.8).

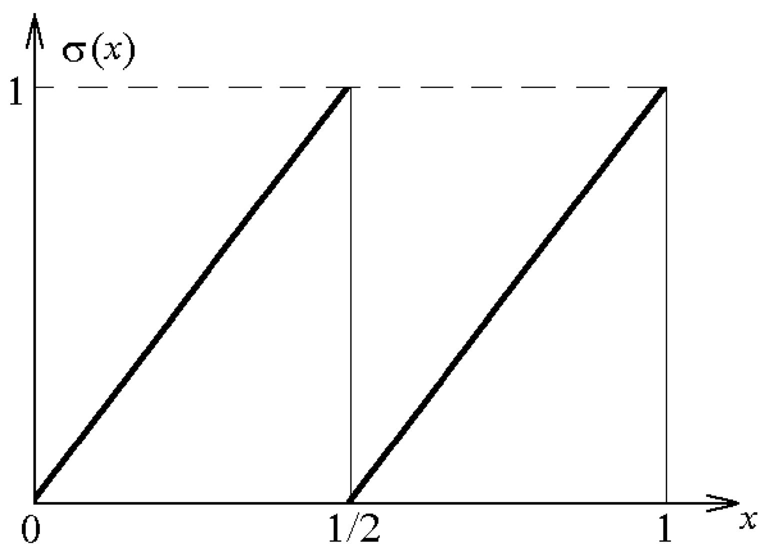


Рис. 5.3. График отображения "сдвиг Бернулли"

1. Если двоичные дроби x_0 и x_0' отличаются лишь в $n+1$-м знаке, то под действием преобразования σ это различие увеличивается от шага к шагу итерации, и через n итераций x_0 и x_0' будут отличаться в первом разряде.

2. Если x_0 — иррациональное число, то последовательность $\{a_n\}$ первых двоичных разрядов чисел $x_0, x_1, \ldots$ непериодична, но любая конечная последовательность $\alpha_1, \ldots, \alpha_k$ встречается в $\{a_n\}$ бесконечное (счетное) число раз. В этом смысле последовательность $\{a_n\}$ "похожа" по своим свойствам на последовательность орлов и решек, полученных при подбрасывании монеты, если выпадение орла сопоставить с единицей, а решки — с нулем.

3. Образ произвольной иррациональной точки $x_0 \in (0,1)$ подходит к любой фиксированной точке $x \in (0,1)$ на расстояние не более $\varepsilon > 0$ бесконечное (счетное) число раз.

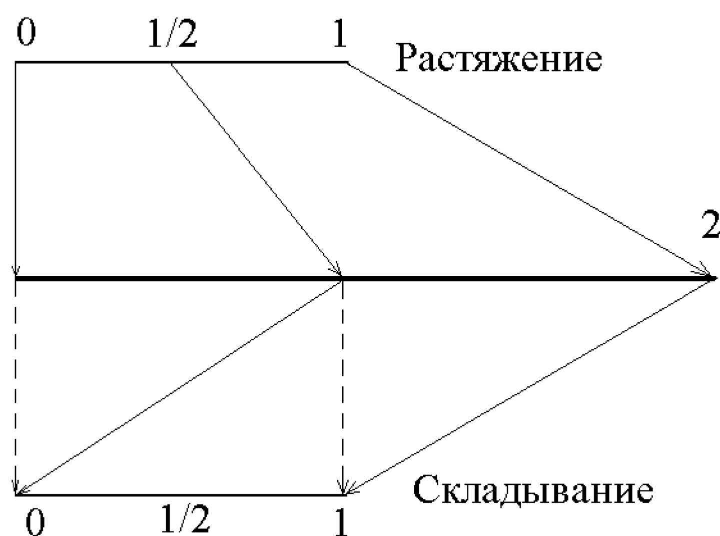


Рис. 5.4. Растяжение и складывание – характерные элементы отображений, приводящих к хаотическим последовательностям

Эти свойства характеризуют последовательность $x_0, x_1, \ldots$ как хаотическую.

Механизм, приводящие к хаосу при сдвигах Бернулли, достаточно универсален. Основные элементы отображения (5.8) — сдвиг и растяжение (рис. 5.4), характерны для большинства хаотических отображений.

Заметим, что одной из причин хаоса является неустойчивость динамической системы, одной из характеристик которой служит показатель Ляпунова, введенный нами ранее для непрерывного случая и характеризующий скорость разбегания фазовых траекторий по сравнению с экспоненциальной. В случае дискретных отображений типа (5.8) показатель Ляпунова

можно определить из схожих соображений: отрезок длины ε после n итераций в случае экспоненциального разбегания фазовых траекторий будет иметь длину порядка $\varepsilon e^{\lambda(x_0)n}$, т.е. $\varepsilon e^{\lambda(x_0)n} = |f^n(x_0+\varepsilon)-f^n(x_0)|$, откуда

$$\begin{aligned}\lambda(x_0) &= \lim_{n\to\infty}\lim_{\varepsilon\to 0}\frac{1}{n}\ln\left|\frac{f^n(x_0+\varepsilon)-f^n(x_0)}{\varepsilon}\right| = \\ &= \lim_{n\to\infty}\frac{1}{n}\ln\left|\frac{df^n(x_0)}{dx_0}\right|;\end{aligned} \tag{5.9}$$

что означает, что $e^{\lambda(x_0)}$ — “среднее” растяжение расстояния между близкими начальными точками за одну итерацию.

Еще одно выражение для показателя Ляпунова может быть получено из (5.9) путем многократного применения формулы для производной сложной функции:

$$\frac{d}{dx}f^2(x)\,|_{x=x_0} = \frac{d}{dx}f(f(x))\,|_{x=x_0} = f'(f(x_0))f'(x_0) = f'(x_1)f'(x_0),$$

где учтено, что $x_1 = f(x_0)$; это дает для $\lambda(x_0)$ формулу

$$\begin{aligned}\lambda(x_0) &= \lim_{n\to\infty}\frac{1}{n}\ln\left|\frac{df^n(x_0)}{dx_0}\right| = \lim_{n\to\infty}\frac{1}{n}\ln\left|\prod_{i=0}^{n-1}f'(x_i)\right| = \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\ln|f'(x_i)|\,.\end{aligned}$$

Эта формула по своей структуре напоминает выражение для энтропии, которая характеризует “неупорядоченность” и количество информации.

5.2.2. Треугольное отображение. Рассмотрим еще один пример дискретного одномерного отображения, приводящего к хаосу — так называемого треугольного отображения. Оно задается соотношением

$$x_{n+1} = \Delta(x_n), \quad n = 1, 2, \ldots, \tag{5.10}$$

где

$$\Delta(x) = r\left(1 - 2\left|\frac{1}{2} - x\right|\right).$$

его график приведен на рис. 5.5.

Исследуем особые точки отображения (5.10), соответствующие точкам “равновесия”; такими точками являются неподвижные точки функции $f(\cdot)$, т.е. такие, для которых выполнено соотношение

$x = f(x)$. На графике отображения $f(\cdot)$ ими являются абсциссы точек пересечения графика функции $f(\cdot)$ с биссектрисой первого координатного угла.

Исследуем устойчивость неподвижной точки x^* отображения $f(\cdot)$. Для этого запишем

$$\delta_{n+1} \equiv |x_{n+1} - x^*| = |f(x_n) - x^*| = |f(x^* + \delta_n) - x^*| \approx \left|\frac{df(x^*)}{dx^*}\right| \delta_n,$$

откуда следует, что критерием устойчивости неподвижной точки x^* является выполнение неравенства $|f'(x^*)| < 1$.

Вернемся к треугольному отображению (5.10) и заметим, что если параметр r удовлетворяет условию $r < 1/2$, то $x = 0$ — единственная неподвижная точка, к которой притягиваются все точки из интервала (0,1); при $r > 1/2$ существуют две неустойчивые неподвижные точки, при этом последовательность $\{x_n\}$ приобретает свойства хаотичности. При $r = 1$ график отображения $\Delta^n(\cdot)$ изображен на рис. 5.6. Видно, что модуль производной отображения $\Delta^n(\cdot)$ равен 2^n, что приводит к значению показателя Ляпунова, равному $\lambda = \ln 2 > 0$.

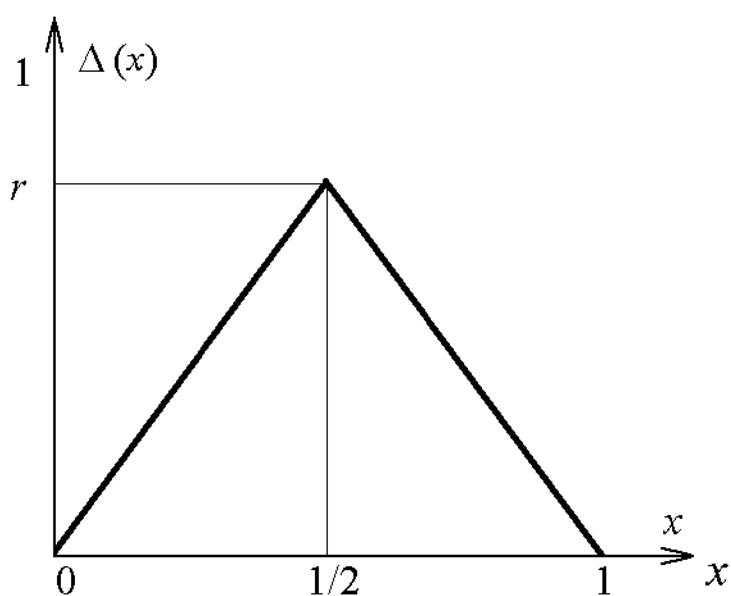


Рис. 5.5. График треугольного отображения

Итак, можно констатировать связь показателя Ляпунова с хаосом: при $1/2 < r \le 1$ показатель Ляпунова $\lambda = \ln 2r > 0$ и в системе наблюдается хаотический режим, при $0 < r < 1/2$ $\lambda = \ln 2r < 0$ и в системе есть притягивающая точка, т.е. хаоса не возникает.

Рассмотрим так называемую выборочную меру на отрезке $[0, 1]$: для любого (борелевского) множества $\mathcal{B} \subset [0, 1]$ обозначим

$$P_N(\mathcal{B}) = \frac{1}{N}\sum_{i=0}^{N-1} \chi_{\mathcal{B}}(x_i),$$

где $\chi_{\mathcal{B}}$ — характеристическая функция множества $\mathcal{B}$:

$$\chi_{\mathcal{B}}(x) = \begin{cases} 1, & \text{если } x \in \mathcal{B}, \\ 0, & \text{если } x \overline{\in} \mathcal{B}. \end{cases}$$

При определенных условиях последовательность мер $\{P_N\}$ сходится к σ-аддитивной мере P на $[0, 1]$, $P([0, 1]) = 1$, т.е. к вероятностной

мере. Формально запишем плотность этой меры:

$$\rho(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \delta(x - f^i(x_0)). \tag{5.11}$$

Если $\{P_n\}$ сходится к вероятностной мере, не зависящей от последовательности $\{x_n\}$, (в таких случаях говорят об инвариантной мере), то усреднение по этой мере можно использовать вместо усреднения по реализациям, (т.е., в некотором смысле — по времени). В этом случае говорят об эргодичности отображения. Итак, для инвариантной меры P и для любой борелевской функции $g(\cdot)$ справедливо

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(x_i) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(f^i(x_0)),$$

что по форме напоминает термодинамическое усреднение в статистической физике, где среднее по времени заменяется на среднее по ансамблю.

Найдем плотность инвариантной меры $\rho(\cdot)$. Заметим, что так как за одну итерацию точка x_0 переходит в $f(x_0)$, то распределение $\delta(x - x_0)$ переходит в распределение $\delta(x - f(x_0))$, и если мы достигли предела в (5.11) на некоторой последовательности $\{x_n\}$, то, сдвинув эту последовательность так, чтобы она начиналась с точки $x_1 = f(x_0)$, в случае существования инвариантной меры мы не изменим предельную плотность $\rho(\cdot)$, т.е. справедливо соотношение

$$\rho(y) = \int_0^1 \delta(y - f(x))\rho(x)dx, \tag{5.12}$$

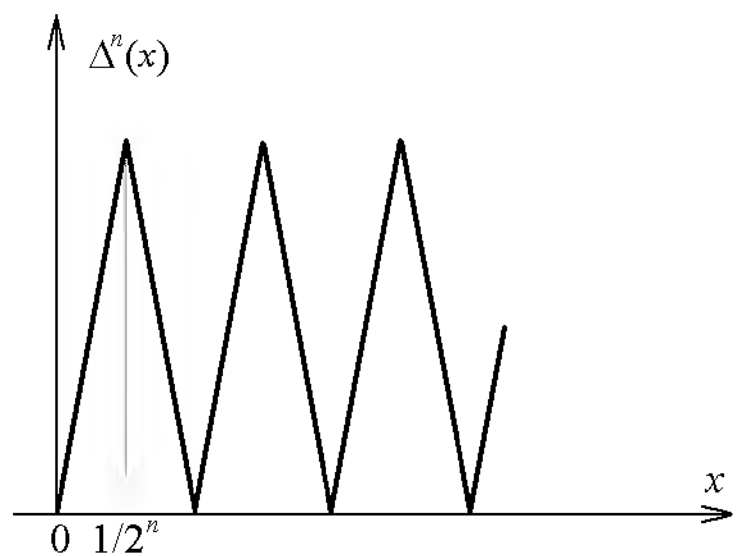


Рис. 5.6. График n-й степени треугольного отображения

это уравнение относительно $\rho(\cdot)$ носит название *уравнения Фробениуса–Перрона*. Для треугольного отображения уравнение (5.12) примет вид

$$\rho(x) = \frac{1}{2}\left(\rho\left(\frac{x}{2}\right) + \rho\left(1 - \frac{x}{2}\right)\right), \tag{5.13}$$

и имеет решение $\rho(x) = 1$. Легко показать, что это — единственное решение (5.13), так как к нему сходится любая последовательность распределений, полученная по формуле

$$\rho_{n+1}(y) = \int_0^1 \delta(y - \Delta(x))\rho_n(x)dx$$

при любом начальном $\rho_0(\cdot)$. Существование такого решения означает, что в пределе при $n \to \infty$ точки $x_k = \Delta^k(x_0)$ равномерно заполняют весь интервал $[0, 1]$.

Еще один важный результат, иллюстрирующий свойства треугольного отображения и возникающий при хаосе, состоит в δ-корре- лированности последовательности $\{x_k\}$. Определим автоковариационную функцию $C(m)$ этой последовательности следующим образом:

$$C(m) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \widehat{x}_{i+m}\widehat{x}_i, \quad m = 0, 1, 2, \ldots,$$

где

$$\widehat{x}_i = f^i(x_0) - \overline{x}, \quad \overline{x} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f^i(x_0).$$

Если воспользоваться тем, что выборочная мера сходится к инвариантной мере на $[0, 1]$, то выражение для нахождения $C(m)$ можно переписать, учитывая возможность усреднения по этой мере. После такой процедуры оно примет вид

$$C(m) = \int_0^1 x f^m(x)\rho(x)dx - \left(\int_0^1 x\rho(x)dx\right)^2,$$

и для треугольного отображения для автокорреляционной функции получим

$$C(m) = \frac{1}{12}\delta_{m,0}.$$

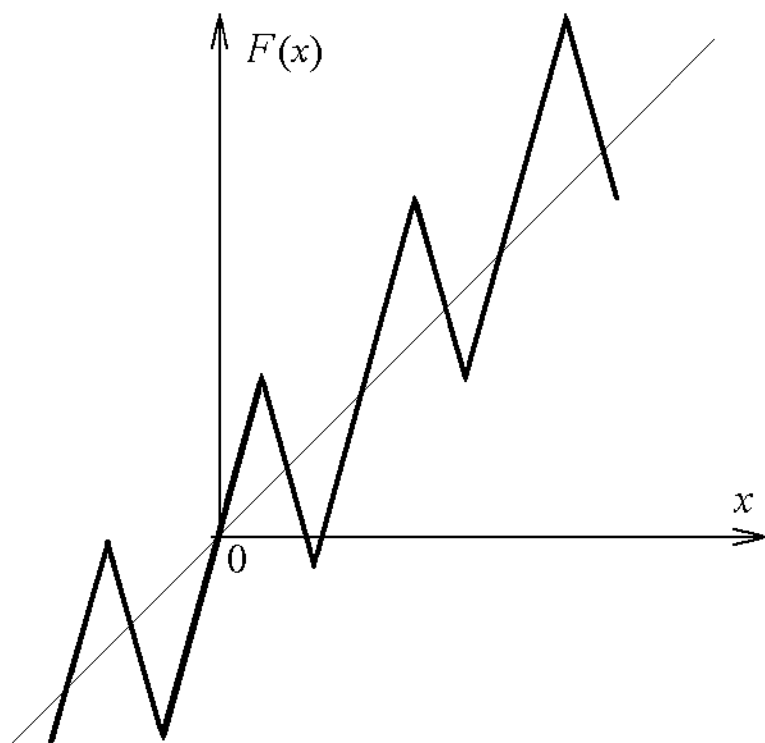


Рис. 5.7. График отображения, приводящего к хаотической диффузии

5.2.3. Математические характеристики хаоса. Обобщая рассмотренные здесь примеры, заключим, что для одномерного отображения вида (5.8) в общем случае последовательность x_0, $f(x_0), \ldots, f^n(x_0), \ldots$ может быть описана с точки зрения:

— показателя Ляпунова, характеризующего устойчивость отображения;

— плотности инвариантной меры, характеризующей то, как распределяются точки данной последовательности на интервале $[0, 1]$;

— автоковариационной функции, характеризующей зависимость между итерациями через m шагов.

5.2.4. Хаотическая диффузия. В заключение этого раздела приведем пример отображения, приводящего к хаотической диффузии в строго детерминированной динамической системе. Это отображение вида

$$x_{n+1} = x + F(x_n), \quad n = 0, 1, 2 \ldots, \tag{5.14}$$

где функция $F(\cdot)$ — периодичная, например, с периодом, равным единице: $F(x+k) = F(x)$, $k = 0, \pm 1, \pm 2, \ldots$ Например, эта функция представляет собой сумму отображения Бернулли, определяющего хаотическое поведение координаты x, и линейной функции, задающей "снос". График рассматриваемого отображения приведен на рис. 5.7. Отображение (5.14) приводит к блужданию точки по всей прямой подобно броуновскому одномерному движению; такое явление носит название детерминированной диффузии.

5.3. Сценарии перехода к хаосу

В этом разделе рассматриваются различные пути перехода к хаосу. Согласно современным представлениям, большинство систем переходит к хаосу по одному из небольшого числа сценариев, что дает надежду на построение достаточно общей качественной теории нелинейных хаотических динамических систем.

Исследование механизмов и способов перехода к хаосу играет важную роль на практике, так как в ряде случаев позволяет предсказать возможность наступления хаотического режима поведения динамической системы при изменении управляющих параметров. Всего будет рассмотрено три сценария, которые к настоящему времени, по-видимому, описывают наиболее осознанные механизмы проявления хаотического поведения динамической системы небольшой размерности — это переход к хаосу через удвоение периода цикла, так называемый переход через перемежаемость, и переход по сценарию Рюэля–Такенса. Сценарии описывают поведение системы на границе регулярного и хаотического режима, т.е. вблизи критического значения управляющего параметра, когда эволюция системы еще не полностью беспорядочна, но уже и не вполне регулярна. Такое подробное изучение окрестности точки перехода к хаосу можно образно назвать "распаковкой" точки бифуркации, разделяющей регулярное и нерегулярное движение системы, рассматриванием критической точки "в микроскоп".

5.3.1. Переход к хаосу через удвоение периода. Рассмотрим так называемое *логистическое отображение*, задаваемое соотношением

$$x_{n+1} = f_r(x_n) \equiv r x_n (1 - x_n), \quad x \in [0, 1]. \tag{5.15}$$

Его график изображен на рис. 5.8.

Впервые это отображение было исследовано П.Ф.Ферхюльстом в 1845 г. для описания роста числа особей популяции в некоторой замкнутой среде. Численность популяции x_{n+1}, согласно модели П.Ф.Ферхюльста, пропорциональна, с одной стороны, количеству особей x_n в предыдущий год, а с другой — площади свободного жизненного пространства. Можно предположить, таким образом, что она пропорциональна их произведению. Но жизненное пространство тем меньше, чем больше численность популяции. Предположив "гладкость" этой зависимости и ограничившись лишь линейными членами в разложении ее по формуле Тейлора, получим, что площадь жизненного пространства прямо пропорциональна величине $1-x_n$, что и дает зависимость (5.15). Другой пример, приводящий к логистическому преобразованию (5.15), связан с моделью банковских сбережений, растущих по закону "сложного процента", т.е. когда размер вклада x_{n+1} в $(n+1)$-й год равен

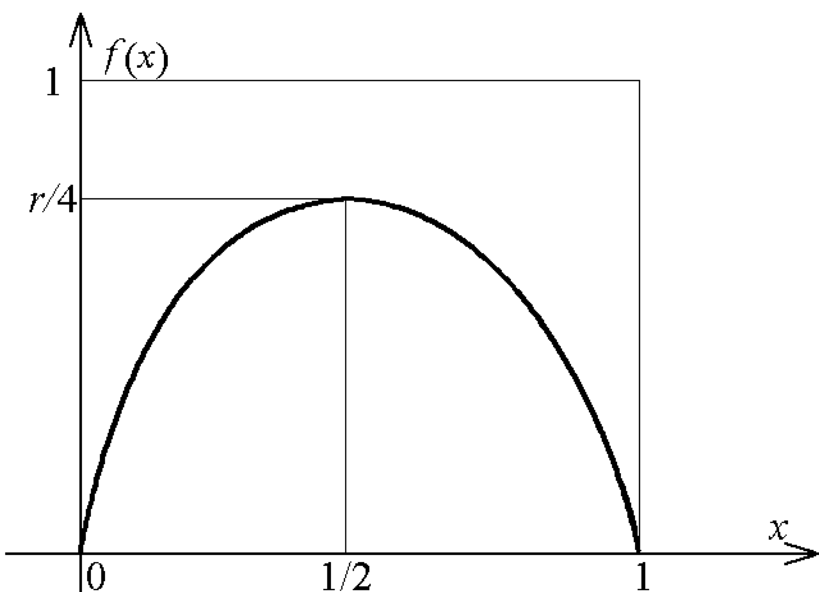


Рис. 5.8. График логистического отображения

$$x_{n+1} = (1+\varepsilon)x_n = \ldots$$

$$\ldots = (1+\varepsilon)^{n+1}x_0,$$

а значит, растет экспоненциально; здесь ε — годовой процент. Владельцу банка, заботящемуся о доходности своего предприятия, такой быстрый рост размера вклада может показаться чрезмерным, и он может ввести максимальный размер вклада $x_{\max}$, а для этого установить переменный годовой процент, заменив ε на $\varepsilon(1 - x_n/x_{\max})$. После очевидных преобразований мы вновь прийдем к логистическому отображению (5.15).

Казалось бы, из соображений "здравого смысла" ситуация с ростом популяции, как и ситуация с ростом размера банковского вклада, должна привести к некоторой стабилизации x_n при $n \to \infty$; так и происходит при некоторых значениях управляющего параметра r, однако существуют и такие значения r, при которых последовательность $x_0, x_1, \ldots, x_n, \ldots$ демонстрирует хаотическое поведение.

Проиллюстрируем это на примере численного исследования логистического отображения, построив бифуркационную диаграмму, рис. 5.9, по оси абсцисс мы будем откладывать значение управляющего параметра, а по оси ординат — положение предельных точек отображения (5.15), т.е. таких точек, для каждой из которых существует подпоследовательность $\{x_{n_k}\}$ последовательности $\{x_n\}$,

сходящаяся к ней.

При анализе приведенного рисунка можно сделать следующие выводы.

1. При некоторых значениях $r = r_n$ управляющего параметра происходит бифуркация, в результате которой каждая предельная точка отображения $f_r(\cdot)$ распадается на две.

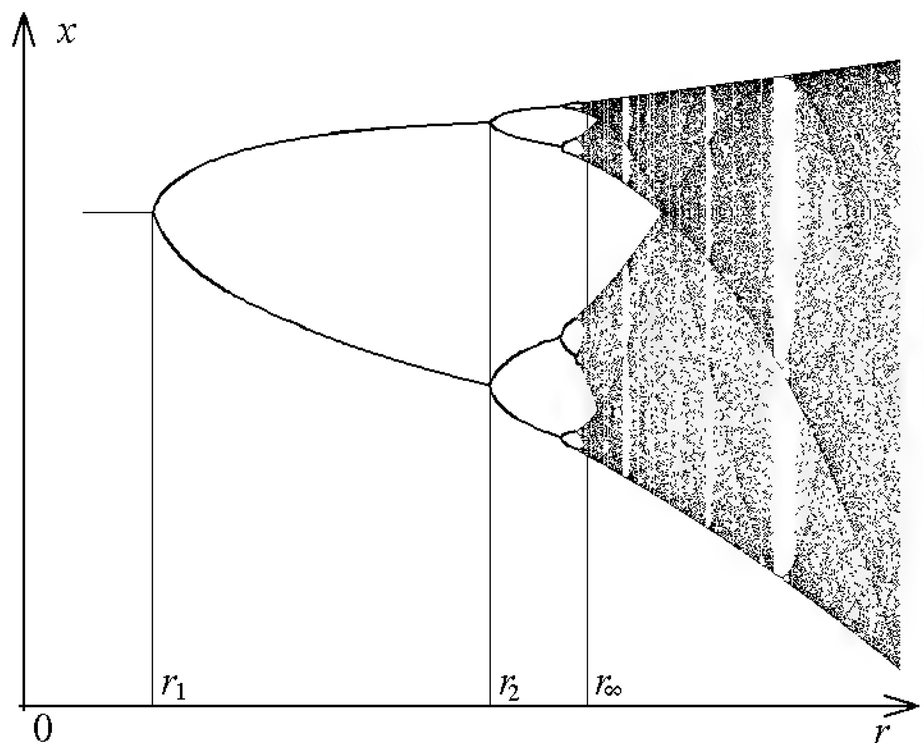


Рис. 5.9. Положение предельных точек логистической последовательности в зависимости от управляющего параметра

2. При n-й бифуркации число предельных точек становится равным 2^n. Значения параметров r_n, при которых происходит удвоение их числа, удовлетворяет масштабному соотношению $r_n = r_\infty - \mathrm{const}\,\delta^{-n}$ или, как говорят, имеет место скейлинг. Здесь r_∞ — предел последовательности $\{r_n\}$ при $n \to \infty$, сходящейся с экспоненциальной скоростью.

3. Расстояние d_n от точки $x = 1/2$ до ближайшей к ней, лежащей на одной с ней "вилке" при фиксированном r, рис. 5.10, подчинено соотношению

$$\frac{d_n}{d_{n+1}} = -\alpha, \quad n \gg 1,$$

где $\alpha = 2.5029\ldots$, $\delta = 4.6692\ldots$

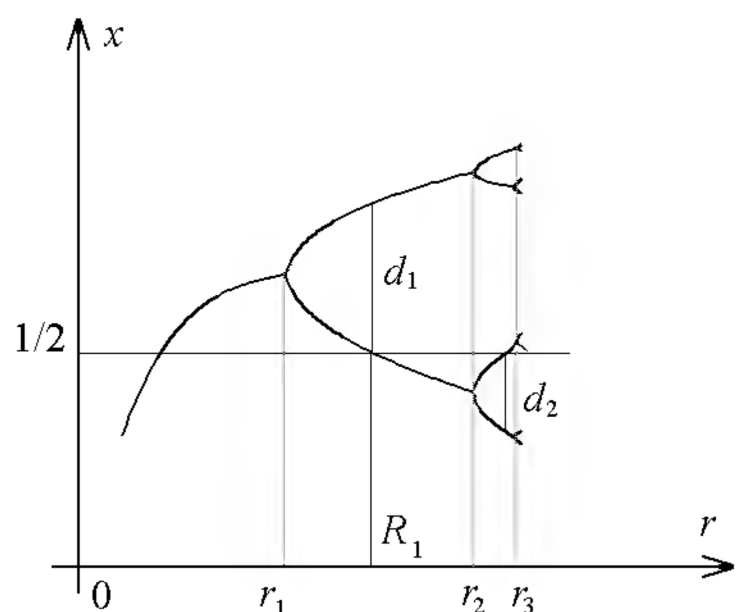


Рис. 5.10. Скейлинг при логистическом отображении

4. При $r > r_\infty$ интервалы значений параметра r, при которых наблюдается сплошное заполнение отрезка $[0, 1]$ предельными точками логистического отображения, изредка сменяются периодами порядка, когда число предельных точек вновь конечно.

Рассмотрим, почему логистическое отображение демонстрирует такое поведение. Для этого проведем исследование неподвижных точек отображения (5.15). Напомним, что графически неподвижные точки изображаются абсциссой точки пересечения графика функции $f_r(\cdot)$ с биссектрисой первого координатного угла. Если $r < 1$, то такая точка — только одна, $x = 0$, и в силу того, что модуль производной $|f_r'(0)| < 1$ при $r < 1$, эта точка — устойчива и является притягивающей, рис. 5.11,*а*.

При $1 < r < 3$ выполняется неравенство $|f_r'(0)| > 1$, что означает, что точка $x = 0$ является неустойчивой, зато появляется еще одна неподвижная точка $x^* = 1 - 1/r$, которая в указанном диапазоне изменения управляющего параметра r оказывается устойчивой, рис. 5.11,*б*.

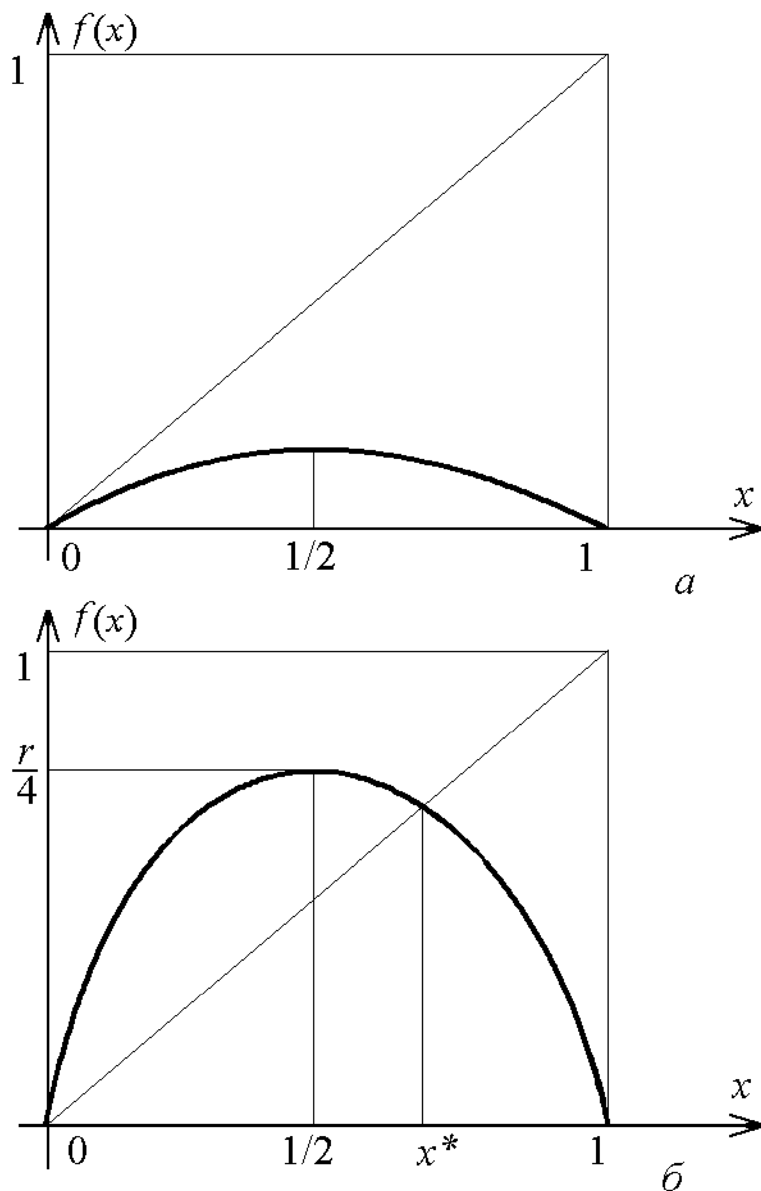


Рис. 5.11. Неподвижные точки логистического отображения

Наконец, если $r > 3$, то модуль производной функции $f_r(\cdot)$ в точке x^* становится больше единицы, что приводит к неустойчивости неподвижной точки x^*. Для того, чтобы понять, что происходит после этого, рассмотрим график функции $f_r^2(\cdot) = f_r(f_r(\cdot))$, приведенный на рис. 5.12. На рис. 5.12,*а* изображен график отображения $f_r(\cdot)$, обладающий двумя неустойчивыми неподвижными точками $x = 0$ и $x = x^*$. На рис. 5.12,*б* приведен график отображения $f_r^2(\cdot)$, у которого, помимо неустойчивых неподвижных точек $x = 0$ и $x = x^*$, есть еще две неподвижные точки $x = x_1$ и $x = x_2$, которые, в отличие от точек $x = 0$ и $x = x^*$, являются устойчивыми, причем $x_2 = f_r(x_1)$, $x_1 = f_r(x_2)$, а значит, $f_r^2(x_i) = x_i$, $i = 1, 2$. На рис. 5.12,*б* в увеличенном масштабе изображена часть графика функции $f_r^2(\cdot)$ в точке пересечения его с биссектрисой первого координатного угла, а на рис. 5.12,*г–е* представлены варианты этого пересечения в зависимости от значения управляющего параметра r. Увеличение числа предельных точек с ростом r связано с тем, что график отображения $f_r^2(\cdot)$ пересекается с биссектрисой первого координатного угла по разному в зависимости от значения r, как это показано на рис. 5.12,*г–е*.

С увеличением r происходит дальнейший рост модуля производной функции $f_r^2(\cdot)$ в точках $x = x_1$ и $x = x_2$, и до тех пор, пока он меньше единицы, точки $x = x_1$ и $x = x_2$ остаются предельными для отображения (5.15). Наконец, при некотором значении $r > r_2$ получается график, изображенный на рис. 5.13. Особенность этого рисунка состоит в том, что части графика, содержащиеся в выделенных квадратах, после растяжения их на весь отрезок $[0, 1]$ становятся “похожими” на график функции $f_r(\cdot)$, и ситуация с удвоением числа предельных точек повторяется.

Обобщая эти результаты, получим следующее.

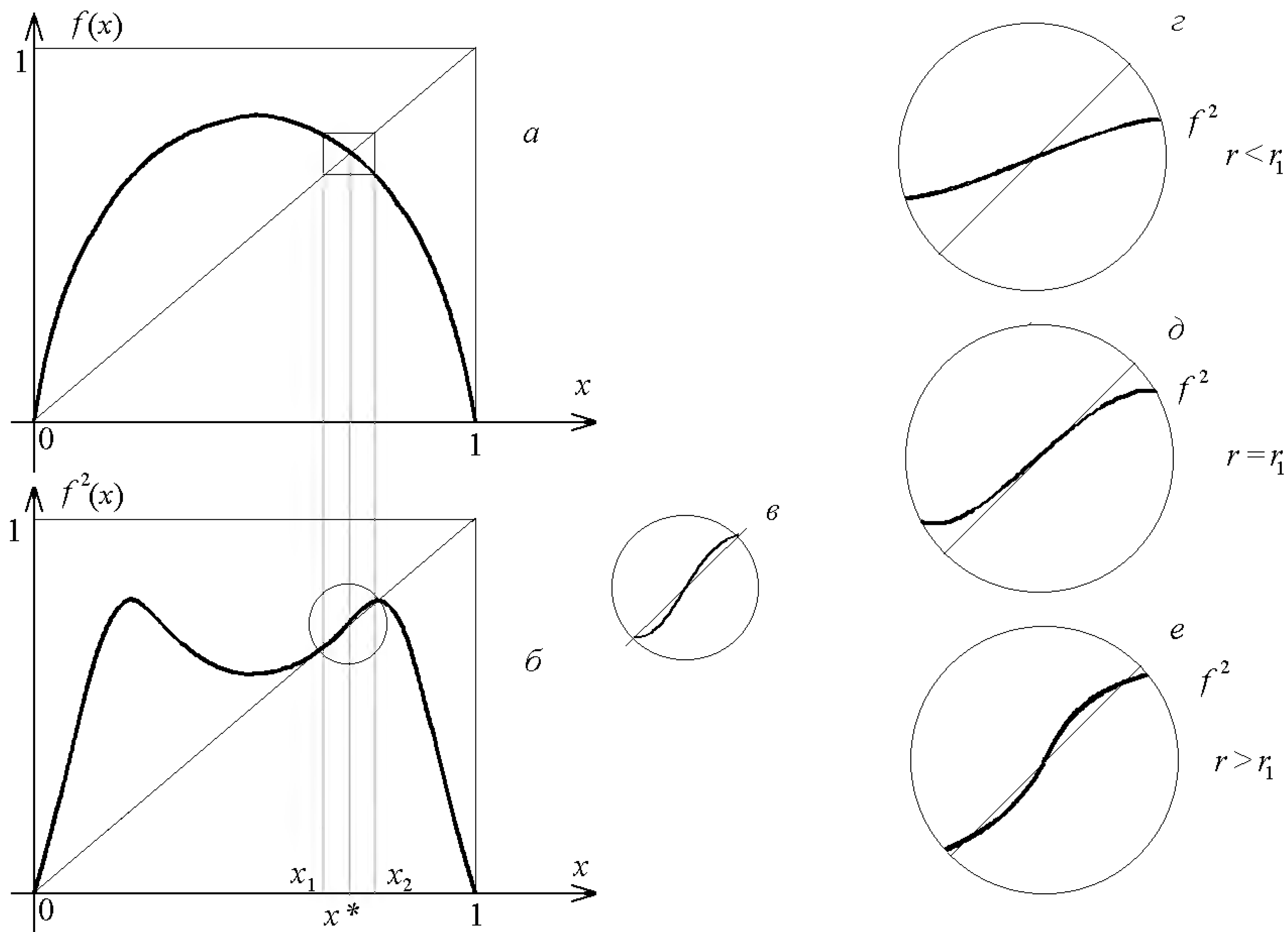


Рис. 5.12. Удвоение периода цикла при логистическом отображении

1. При $r_{n-1} < r < r_n$ имеется устойчивый 2^{n-1}-цикл, т.е. последовательность из 2^{n-1} точек $x_0^*, x_1^*, \ldots, x_{2^{n-1}-1}^*$, такая, что

$$f_r(x_i^*) = x_{i+1}^*, \quad f_r^{2^{n-1}}(x_i^*) = x_i^*,$$

$$\left|\frac{d}{dx} f_r^{2^{n-1}}(x_0^*)\right| = \left|\prod_i f_r'(x_i^*)\right| < 1.$$

2. При $r = r_n$ у отображения $f_r^{2^n}$ происходит бифуркация удвоения числа предельных точек, при которой все точки 2^{n-1}-цикла становятся неустойчивыми, и появляется устойчивый 2^n-цикл.

Оказывается, что это свойство достаточно универсально и выполняется для всех отображений с локально квадратичным максимумом, точнее, для функций с отрицательной на всем интервале $[0,1]$ производной Шварца:

$$Sf \equiv \frac{f'''}{f'} - \frac{3}{2}\left(\frac{f''}{f'}\right)^2 \sim \frac{d^2}{dx^2}[f'(x))]^{-1/2},$$

что выполнено, в частности, для логистического отображения, поскольку $f_r''' = 0$. Это связано с тем, что при многократном масштабном преобразовании центральной части графика функции $f_r(\cdot)$

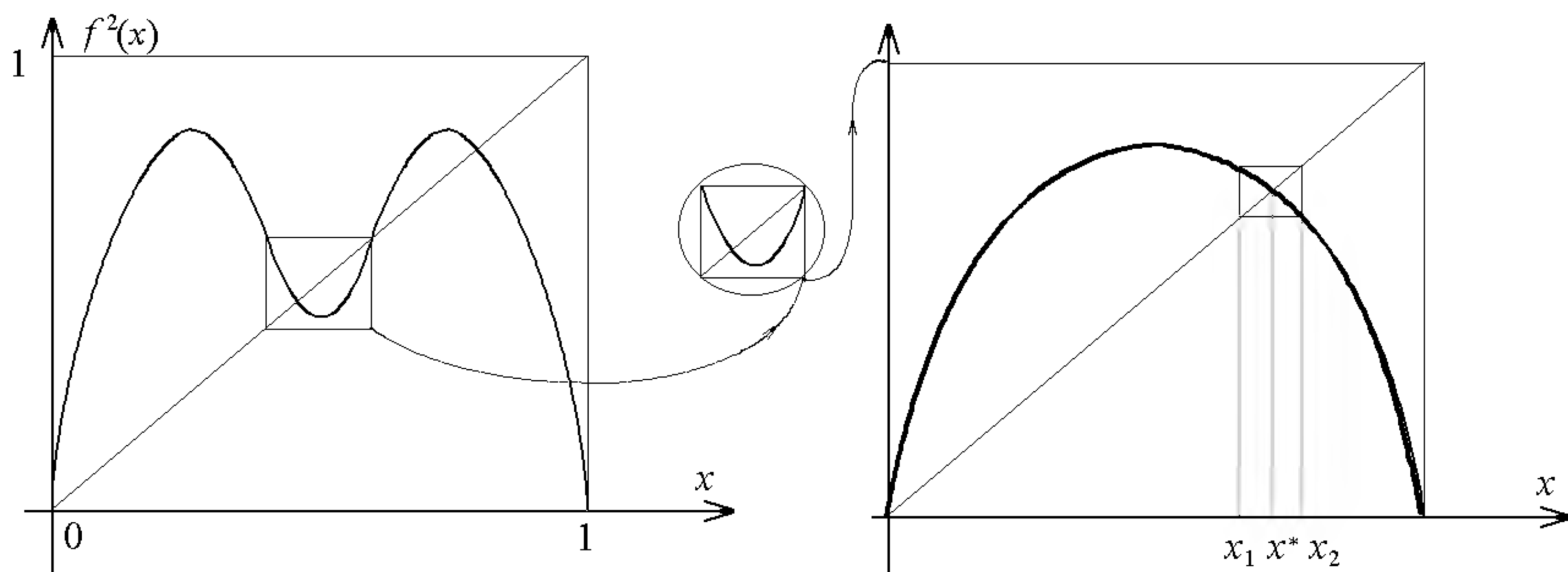


Рис. 5.13. Самоподобие логистического отображения

основную роль в поведении функции $f_r^{2^n}$ начинает играть лишь окрестность максимума функции $f_r(\cdot)$.

Заметим, что показатели Ляпунова при тех значениях параметра r, при которых существуют устойчивые предельные точки, отрицательны, обращаются в нуль в точках бифуркации и положительны при значениях управляющего параметра r, при которых реализуется хаотическое поведение последовательности $\{x_n\}$.

5.3.2. Переход к хаосу через перемежаемость. Сценарий перехода к хаосу через перемежаемость. характерен тем, что при некоторых значениях управляющего параметра $r < r_c$ система демонстрирует регулярное поведение (например, колебания), затем, при превышении критического значения параметра r_c в поведении системы после некоторого периода регулярного движения встречаются хаотические всплески, далее, с ростом значения r, таких всплесков становится все больше, и наконец, движение полностью хаотизируется, рис. 5.14.

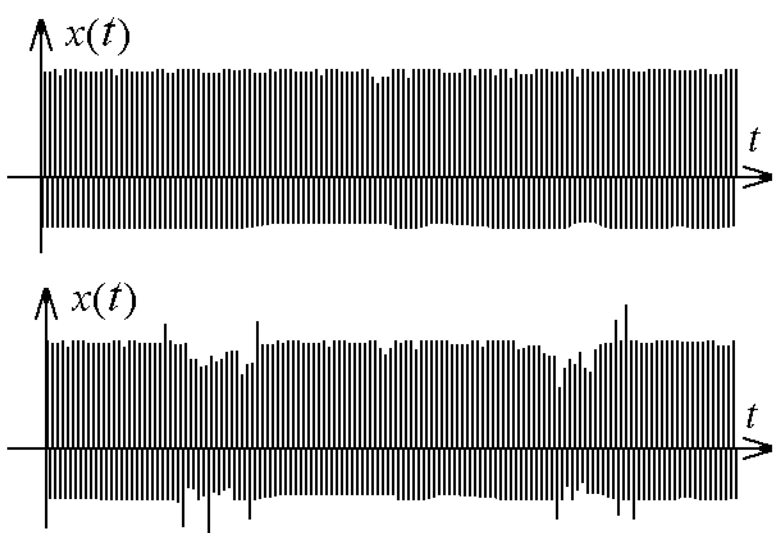


Рис. 5.14. Характерное поведение системы при переходе к хаосу через перемежаемость

На практике такой сценарий перехода к хаосу встречается при конвекции Бенара, в частности, при анализе поведения Y-компоненты координат (X, Y, Z) динамической системы, описываемой уравнениями аттрактора Лоренца.

Отображение Пуанкаре для модели Лоренца при значении параметра, немного превосходящем критическое, изображен на рис. 5.15.

На этом же рисунке изображен процесс перехода графика отображения Пуанкаре через биссектрису первого координатного угла, который демонстрирует слияние (при $\mu = \mu_{\text{кр}}$) двух неподвижных точек — устойчивой и неустойчивой — в одну; при $\mu > \mu_{\text{кр}}$ неподвижных точек не существует. При малом отличии параметра μ от

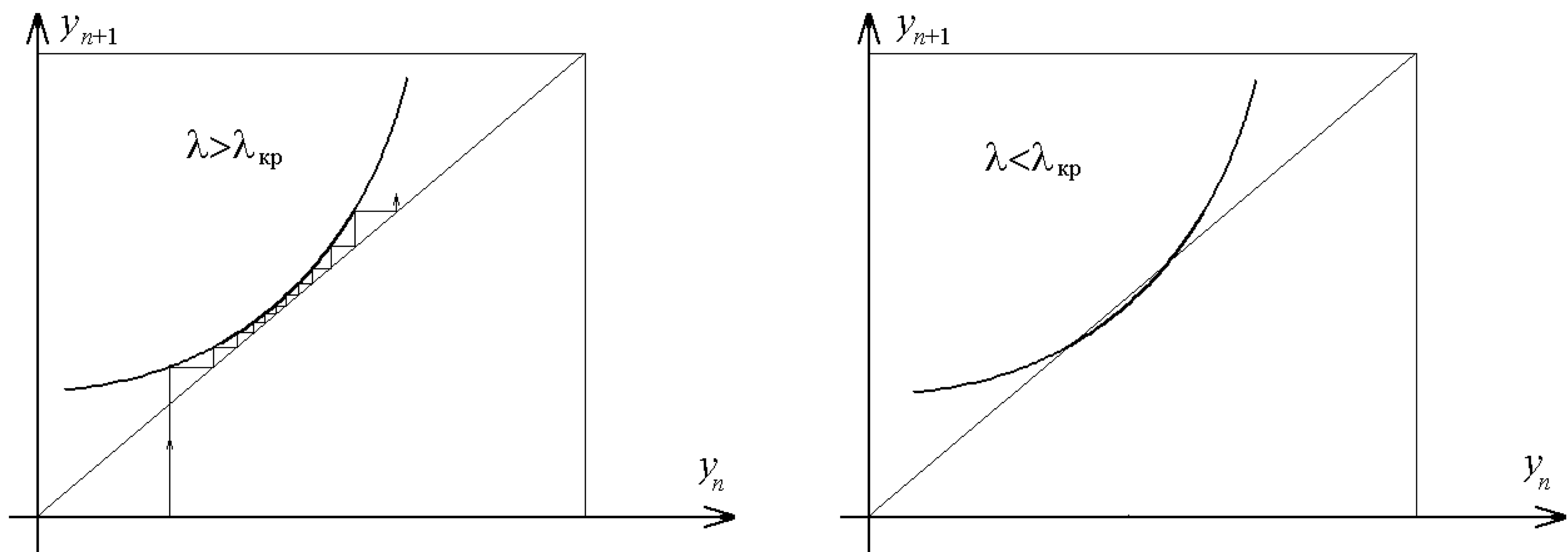


Рис. 5.15. "Просачивание" фазовой траектории между графиком отображения Пуанкаре и координатой первого координатного угла при переходе к хаосу через перемежаемость

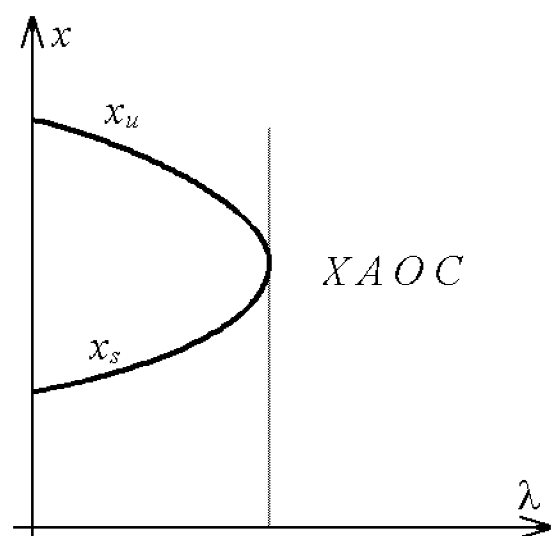


Рис. 5.16. Диаграмма при касательной бифуркации

$\mu_{\text{кр}}$, когда график отображения “почти касается” биссектрисы, образ исчезнувшей неподвижной точки притягивает точки справа и отталкивает слева, при этом “просачивание” точки через узкий участок между графиками отображения $y = f(x)$ и биссектрисы $y = x$ происходит за достаточно большое число итераций, демонстрируя квазирегулярное поведение системы. Возвращение точки в левую, притягивающую часть обеспечивается нелинейностью системы и происходит в соответствии с теоремой Пуанкаре о возвращении. Бифуркационная диаграмма такой системы приведена на рис. 5.16, это так называемая касательная бифуркация. Переход к хаосу через перемежаемость с касательной бифуркацией демонстрирует и логистическое отображение (5.15). При $r = 1 + \sqrt{8}$ это отображение порождает цикл периода 3; его возникновение иллюстрируется графиком, приведенном на рис. 5.17.

Отличие от описанного выше состоит лишь в том, что здесь переход к хаосу осуществляется не при увеличении параметра r, а при уменьшении его. При $r = r_c$ имеется три касания графика функции $f_r^3(\cdot)$ с биссектрисой первого координатного угла, которые при уменьшении управляющего параметра порождают переход к хаосу через перемежаемость после касательной бифуркации, демонстрирующуюся в синхронных возмущениях всех трех неподвижных точек, см. рис. 5.17.

Два другие типа перехода к хаосу через перемежаемость, несколько отличающихся от перехода с касательной бифуркацией, иллюстрируются на рис. 5.18.

На этих рисунках устойчивые неподвижные точки при $r < r_c$, изображенные слева, переходят в неустойчивые при $r > r_c$, изображенные справа. В обеих примерах возврат в окрестность нуля

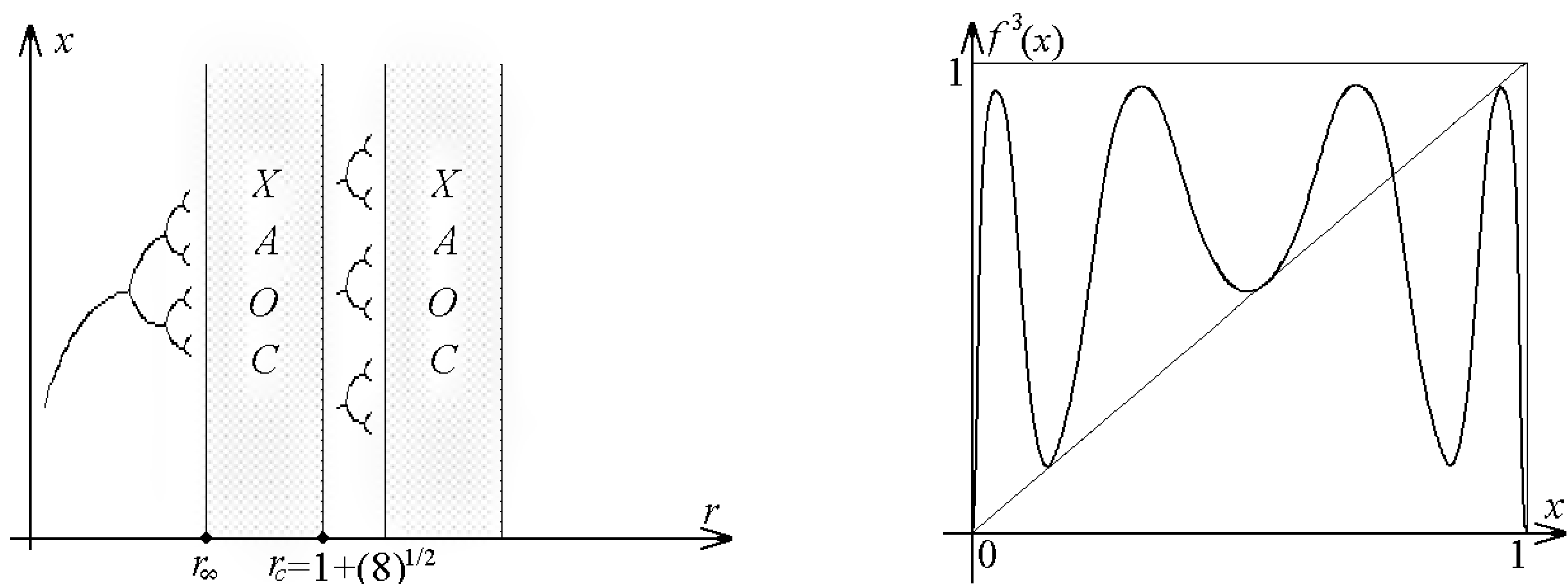


Рис. 5. 17. Переход к хаосу через перемежаемость при логистическом отображении

в соответствии с теоремой Пуанкаре о возвращении происходит за счет нелинейности через некоторое случайное время, после чего точка "просачивается" между графиками функций $y = f(x)$ и $y = x$ и демонстрирует квазирегулярное поведение.

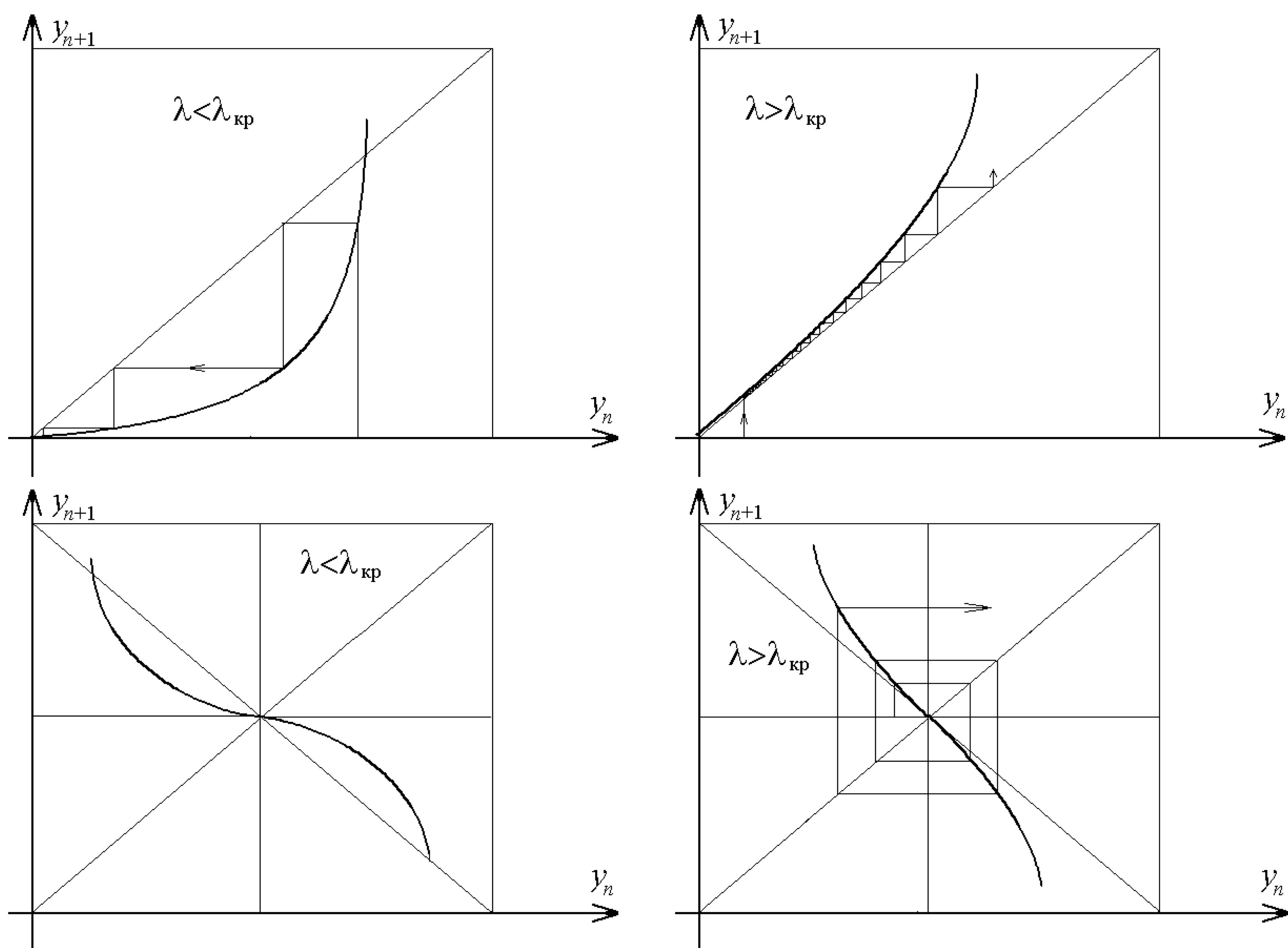


Рис. 5.18. Различные варианты отображений Пуанкаре при переходе к хаосу через перемежаемость

Такие сценарии перехода к хаосу характерны для экспериментов типа химической реакции Белоусова–Жаботинского.

5.3.3. Сценарий Рюэля–Такенса. Рассмотрим кратко еще один способ перехода к хаосу, так называемый сценарий Рюэля–

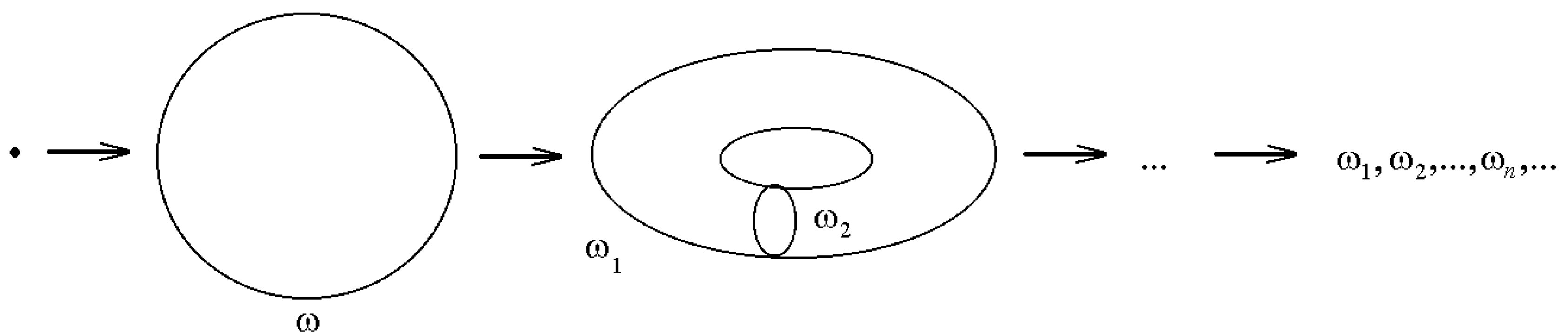


Рис. 5.19. Сценарий развития турбулентности по Ландау

Такенса. Этот сценарий возник из попыток объяснить возникновение турбулентных движений, и опирается на понятие бифуркации Хопфа рождения предельного цикла из устойчивого фокуса, рассмотренного нами ранее, см. рис. 2.13. Бифуркация Андронова–Хопфа приводит к появлению в системе новой частоты, отсутствовавшей в ней ранее. На этом соображении основана модель Ландау возникновения турбулентности, который представлял этот процесс как последовательность бифуркаций Хопфа, и считал, что в начальные этапы развития турбулентности движение системы происходит по инвариантному тору с n несоизмеримыми частотами, $n \to \infty$ (рис. 5.19).

Однако, как показал опыт (в частности, возникновение хаоса в опытах с конвекцией Бенара), хаос наступает сразу после появления двух основных частот. Этот факт стимулировал теоретические исследования, в результате которых Рюэль и Такенс показали, что после двух бифуркаций Хопфа регулярное движение может стать неустойчивым и перейти в хаотическое — так называемое движение на странном аттракторе.

Рассмотренные три сценария, по-видимому, не исчерпывают все возможности перехода к хаосу; этот перечень можно пополнить, например, рассмотрением многомерного управляющего параметра.

5.4. Эргодичность и перемешивание

В предыдущих разделах на примерах одномерных отображений Пуанкаре было показано, что важными характеристиками хаоса является существование инвариантной меры в фазовом пространстве, характеризующей частоту попадания фазовой точки в заданный объем с течением времени, и свойство перемешивания, выражающее запутывание фазовой траектории. Вернемся к этому вопросу с более общих позиций.

5.4.1. Среднее по времени и среднее по ансамблю. Рассмотрим консервативную динамическую систему. Пусть фазовые переменные $\mathbf{x}(t)$ описывают ее состояние в момент времени t, и

$h(\mathbf{x}) = h(\mathbf{x}(t))$ — интегрируемая функция, зависящая от фазовых координат.

Определение. *Средним по времени* значением функции $h(\mathbf{x}(t))$ называется величина

$$\overline{h}(x_0) = \lim_{T\to\infty} \frac{1}{T} \int\limits_0^T h(\mathbf{x}(t))dt.$$

Обозначив $F^t(\cdot)$ фазовый поток, найдем, что

$$\overline{h}(x_0) = \lim_{T\to\infty} \frac{1}{T} \int\limits_0^T h(F^t\mathbf{x}_0)dt,$$

то есть, вообще говоря, среднее значение h по времени зависит от начального состояния x_0. Однако оказывается, что для почти всех (т.е. кроме множества меры нуль) начальных условий функция $\overline{h}(x_0)$ не меняется при перенесении начального условия x_0 вдоль траектории.

Пусть движение системы в фазовом пространстве $\mathcal{R}_n$ происходит в ограниченной области $D \in \mathcal{R}_n$ с объемом V_D.

Определение. *Фазовым средним* значением функции $h(\mathbf{x}(t))$ называется величина

$$\langle h\rangle = \frac{1}{V_D} \int\limits_D h(\mathbf{x})dV, \quad dV \equiv dx_1 \ldots dx_n.$$

5.4.2. Эргодические системы. Движение системы называется эргодическим, а сама система — эргодической, если для любой интегрируемой функции $h(\mathbf{x})$ и для любых начальных условий выполнено равенство

$$\langle h\rangle = \overline{h}(x_0). \tag{5.16}$$

Рассмотрим область $\Omega \in D$, и пусть $h_\Omega(\mathbf{x})$ — характеристическая функция множества Ω. Ее фазовое среднее

$$\langle h\rangle = \frac{1}{V_D} \int\limits_D h_\Omega(\mathbf{x})dV = \frac{V_\Omega}{V_D}$$

равно относительному объему части фазовой области D, занимаемой подобластью $\Omega \in D$. Значит, для эргодической системы относительное время, проводимого системой в Ω, пропорционально относительному объему V_Ω/V_D и не зависит от начального состояния x_0.

Иными словами, фазовая кривая эргодической системы равномерно и плотно заполняет весь объем D.

Для автономной гамильтоновой системы движение происходит по фазовому многообразию, размерность которого меньше, чем размерность всего фазового пространства, так как в таких системах есть первый интеграл — энергия. Поэтому поток может быть эргодичным только на этой поверхности, а если имеются и другие k первых интегралов, то эволюция системы происходит на $(n-k-1)$-мерной гиперповерхности в $\mathcal{R}_n$.

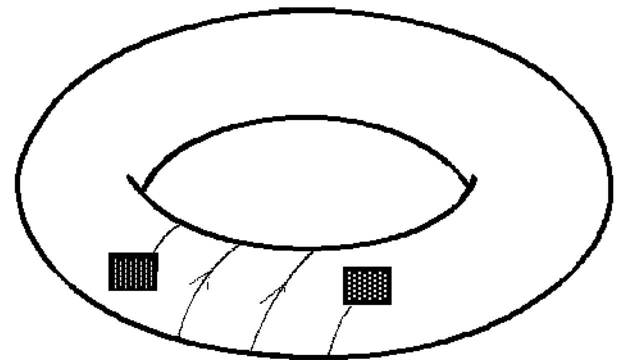


Рис. 5.20. Движение по инвариантному тору эргодичное, но не перемешивающее

Примером эргодичной гамильтоновой системы является движение двух гармонических осцилляторов на двумерном торе с иррациональным отношением частот. В этом случае, как было отмечено в предыдущих разделах, фазовая кривая равномерно и плотно заполняет весь двумерный тор.

5.4.3. Перемешивающие системы. Эргодичность — необходимое, но не достаточное условие хаотичности движения динамической системы, так как, например, в рассматриваемом примере гамильтоновой системы выделенная начальная область Ω целиком перемещается по тору фазовым потоком без изменения ее формы, рис. 5.20. Вместе с тем существуют системы с более сложными режимами движения, для которых начальный объем, перемещаясь по доступной части фазового пространства, сильно деформируется, так, что с течением времени разные части исходной области Ω_0 можно обнаружить в разных частях объема D вне зависимости от формы и расположения Ω_0. Это свойство, называемое перемешиванием, и может служить критерием хаотичности.

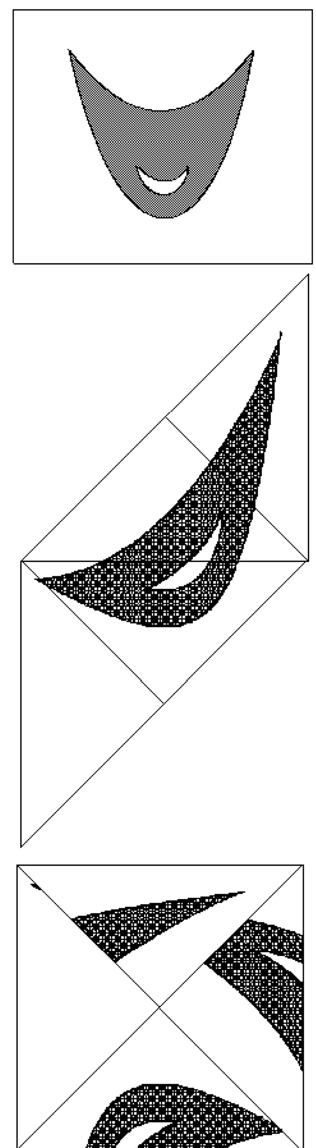


Рис. 5.21. Пример перемешивающего отображения плоскости "Кот Арнольда"

Один из примеров такой системы изображен на рис. 5.21 и носит название “Кота Арнольда”. Преобразование состоит в вытягивании квадрата с изображением улыбающегося кота и последующим сложением из вытянутых частей нового квадрата, и по своей сути является двумерным обобщением сдвига Бернулли.

Другой пример системы с перемешиванием, более близкий к реальным физическим системам, принадлежит Гиббсу. Он состоит в том, что в начальный момент рассматривается сосуд, на треть заполненный чернилами, а на две трети — водой, причем эти жидкости не перемешаны. После встряхивания

сосуда любая часть его объема будет на треть состоять из чернил, и на две трети — из воды. Формализуем эти примеры следующим образом.

Рассмотрим две подобласти A и B области D, и пусть эволюция A описывается фазовым потоком F^t динамической системы так, что $A_t = F^t A$, а область B неподвижна. Обозначим $A_t \cap B$ совокупность всех "частей" A, попавших в B в момент времени t.

Определение. Динамическая система называется *перемешивающей*, а фазовый поток — *перемешивающим*, если при $t \to \infty$ существует предел отношения меры (объема) $\mu(A_t \cap B)$ области $A_t \cap B$ к мере (объему) $\mu(B)$ области B, равный

$$\lim_{t\to\infty} \frac{\mu(A_t \cap B)}{\mu(B)} = \frac{\mu(A)}{\mu(D)}. \tag{5.17}$$

Иными словами, так как $\mu(A)/\mu(D)$ есть относительная часть объема области D, занимаемая областью A в начальный момент времени, то, в силу консервативности динамической системы, по теореме Лиувилля $\mu(A) = \mu(A_t)$, и $(\mu(A_t \cap B))/\mu(B)$ есть относительная часть объема B, занимаемая попавшими туда к моменту времени t "частями" области A. Соотношение (5.17) означает, что в пределе эта относительная часть объема сохраняется независимо от формы, величины и взаимного расположения областей A и B.

Из определения перемешивающих систем можно сделать ряд выводов о характере поведения таких систем. Во-первых, область B можно сделать сколь угодно малой и поместить в любое место области D. Если выполнено (5.17), то это означает, что с ростом t части области A можно найти с одинаковым успехом в любой точке фазового пространства, доступного для движения системы. Следовательно, A с течением времени превращается в сколь угодно тонкую паутинку, пронизывающую всю область D фазового пространства, причем объем этой паутинки равен объему исходной области A. Надо сказать, что топологические свойства таких областей могут оказаться весьма экзотичными, в частности, может потребоваться уточнение понятия объема и размерности для предела A_t при $t \to \infty$.

Второе замечание касается непредсказуемости поведения системы. Так как сколь угодно малая область фазового пространства с течением времени переходит в паутинку, о которой шла речь в предыдущем абзаце, то, в силу невозможности задания начальных условий реальной физической системы с абсолютной точностью, сколь угодно малая область неопределенности начальных условий при достаточно больших временах наблюдения заполнит всю доступную область D фазового пространства, т.е. при достаточно большом фиксированном времени наблюдения t для любой точки области D фазового пространства найдется сколь угодно малая окрестность, содержащая фазовую

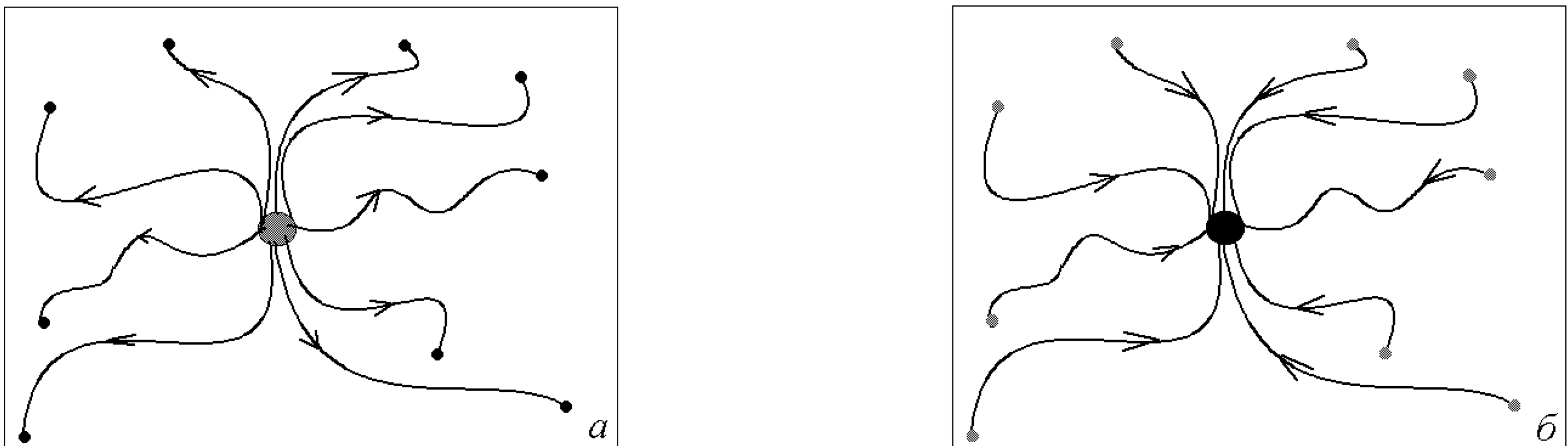


Рис. 5.22. Фазовые траектории: разбегающиеся из окрестности некоторой точки фазового пространства (*а*); попадающие в эту окрестность (*б*)

траекторию, начинающуюся в малой окрестности начальной точки, рис. 5.22,*а*.

Наконец, заметим, что перемешивание приводит к необратимости. Действительно, так как любая неопределенность начальных условий через некоторое время заполняет всю область D, то зная, что в момент времени t система находится в некоторой сколь угодно малой окрестности ε точки $\mathbf{x}(t)$ фазового пространства, мы не можем указать начальное состояние системы, так как к моменту времени t в окрестность ε может попасть траектория, начинающаяся в малой окрестности любой начальной точки, см. рис. 5.22,*б*.

Таком образом, поскольку математическое моделирование физических систем можно производить лишь с конечной точностью, то для систем с перемешиванием детерминированное описание оказывается лишенным смысла, и более адекватным может оказаться стохастическое описание.

Как можно было заметить из примеров, перемешивание влечет эргодичность, и это утверждение может быть строго доказано. Обратное же, вообще говоря, не верно, что демонстрирует ранее рассмотренный пример с условно периодическим движением по двумерному тору.

Заметим, что эффект перемешивания вовсе не связан с большим числом степеней свободы, как это может показаться, если обратиться к статистической механике, в которой эффект перемешивания наблюдается в системе классического идеального газа. В 1962 г. появилась работа Я.Г. Синая, в которой доказывалось, что система из двух абсолютно упругих дисков, движущихся в ящике с упругими стенками, обладает эффектом перемешивания (биллиард Синая). Позже было показано, что теми же свойствами обладает и система с одним движущимся диском и одним закрепленным (но обладающим вдвое большим диаметром), т.е. система с еще меньшим числом степеней свободы.

5.4.4. Диссипативные перемешивающие системы. Рассмотрим теперь диссипативные системы, у которых фазовый объем может убывать со временем. Сокращение фазового объема приводит

к тому, что при $n \to \infty$ состояние диссипативной динамической системы описывается фазовой точкой, стремящейся к некоторому подмножеству фазового пространства, называемому аттрактором. В зависимости от того, является ли аттрактор простым или сложным, движение системы может быть регулярным или хаотическим.

Более строго, если система описывается эволюционным уравнением

$$\dot{\mathbf{x}} = F(\mathbf{x}), \tag{5.18}$$

то убывание фазового объема происходит тогда, когда поле фазовых скоростей имеет отрицательную дивергенцию: $\operatorname{div} F(\mathbf{x}) < 0$, в этом случае существует подобласть B фазового пространства, называемая аттрактором, к которому сходятся при $t \to \infty$ фазовые траектории из некоторой окрестности B.

Точнее, B называется аттрактором, если:

— множество B инвариантно относительно фазового потока;

— существует окрестность U множества B фазового пространства, сжимающаяся к B при $t \to \infty$;

— B не представимо в виде суммы двух своих непересекающихся подмножеств, инвариантных относительно фазового потока.

К простым аттракторам относятся фокусы и предельные циклы, и этим исчерпываются возможности для двумерных динамических систем. Однако если число фазовых переменных больше двух, то возможно существование более сложных аттракторов, к которым в первую очередь относятся многомерные торы с несоизмеримыми частотами, описывающие условно периодические движения. Эти аттракторы, хотя и обладают свойствами эргодичности, не приводят к хаотическим режимам эволюции динамических систем. Характерное свойство простых аттракторов состоит в том, что все они являются так называемыми подмногообразиями фазового пространства $\mathcal{R}$, т.е. такими множествами W из $\mathcal{R}$, для которых выполнены следующие два условия:

— множество W целиком принадлежит пространству $\mathcal{R}' \subseteq \mathcal{R}$;

— в каждой точке $W \subset \mathcal{R}'$ существует единственная касательная гиперплоскость из $\mathcal{R}'$.

Иными словами, W вложено в $\mathcal{R}$ гладко.

5.4.5. Странные аттракторы. Интересными и необычными свойствами обладают так называемые странные аттракторы, возникающие при числе фазовых координат $n \geq 3$ и приводящие к хаотическим режимам эволюции системы. Это такие ограниченные притягивающие множества, которые являются аттракторами и не являются подмногообразиями. Структура этих множеств весьма сложна, а топологические свойства — экзотичны. В частности, для них оказывается достаточно бесполезными обычные понятия объема (длины) или топологической размерности.

Сжатие фазового объема диссипативной динамической системы приводит к тому, что фазовые кривые стягиваются к аттрактору, и попав в некоторую область фазового пространства, остаются в ней навсегда. На самом же аттракторе в случае хаотического режима движение неустойчиво, и две близких фазовых траектории расходятся экспоненциально быстро, но остаются на аттракторе. Таким образом, динамическая система со странным аттрактором характеризуется глобальным сжатием фазового объема и локальной неустойчивостью фазовых траекторий.

Примером странного аттрактора служит аттрактор Лоренца, рассмотренный нами ранее в связи с исследованием конвекции Бенара.

Важной проблемой теории диссипативных динамических систем является нахождение их описания, позволяющего отличать регулярное поведение динамической системы от хаотического.

Рассмотрим при $t = 0$ две начальных точки, $\mathbf{x}_1(0)$ и $\mathbf{x}_2(0)$, и выпустим из них две фазовые траектории. Если режим — хаотический, то расстояние $d(t)$ между этими фазовыми траекториями с течением времени будет экспоненциально увеличиваться, т.е.

$$d(t) \approx d(0)e^{ht}.$$

Средняя скорость этого разбегания дается формулой

$$h \approx \frac{1}{t} \ln \frac{d(t)}{d(0)}. \tag{5.19}$$

Но так как движение финитно, то при больших t, вне зависимости от того, является движение регулярным или хаотическим, $d(t) \leq M$, и значит, $h \to 0$ при $t \to \infty$. Для характеристики движения будем рассматривать лишь локальное поведение траекторий, для чего в (5.19) устремим одновременно t к бесконечности и $d(0)$ к нулю, в результате чего получим

$$h = \lim_{\substack{d(0)\to 0\\ t\to\infty}} \frac{1}{t} \ln \frac{d(t)}{d(0)}.$$

Величина h носит название энтропии Колмогорова–Синая, и от нее зависит характер движения системы. Например, при периодическом или квазипериодическом движении $h = 0$, если система находится в окрестности устойчивой предельной точки, то $h < 0$, а в случае хаотического движения всегда $h > 0$.

Заметим, что так определенная энтропия h может зависеть от направления вектора $\mathbf{x}_1(0) - \mathbf{x}_2(0)$ в фазовом пространстве. Действительно, в разделе 2.3.5 мы говорили, что экспоненциальное разбегание траекторий описывают характеристические показатели Ляпунова,

причем скорости разбегания различны в зависимости от направления $\mathbf{x}_1(0) - \mathbf{x}_2(0)$: если это направление совпадает с направлением сингулярного[1] вектора матрицы линеаризации правой части системы (5.18), то показатель Ляпунова совпадает с h и равен реальной части сингулярного (собственного) значения матрицы линеаризации, соответствующего этому вектору. Если же направление $\mathbf{x}_1(0) - \mathbf{x}_2(0)$ выбрано произвольно, то энтропия Колмогорова–Синая равна максимальному значению реальных частей сингулярных (собственных) значений, соответствующих тем сингулярным векторам, которые не ортогональны вектору $\mathbf{x}_1(0) - \mathbf{x}_2(0)$, дающему максимальную скорость экспоненциального роста $d(t)$ в заданном направлении. Это означает, что энтропия Колмогорова–Синая равна максимальному характеристическому показателю Ляпунова почти всегда, т.е. тогда, когда вектор $\mathbf{x}_1(0) - \mathbf{x}_2(0)$ не ортогонален сингулярному вектору матрицы линеаризации.

Набор характеристических показателей Ляпунова является более тонкой характеристикой динамической системы. Сигнатура показателей Ляпунова позволяет дать классификацию диссипативных динамических систем. Действительно, для одномерных динамических систем возможно лишь отрицательное значение показателя Ляпунова. При $n = 2$ аттрактором может быть либо устойчивый фокус, что соответствует сигнатуре характеристических показателей Ляпунова $(-,-)$, либо предельный цикл с сигнатурой $(0,-)$.

Таблица 5.1

−	−	−	Устойчивый фокус
0	−	−	Устойчивый предельный цикл
0	0	−	Устойчивый тор
+	0	−	Странный аттрактор

В трехмерном случае набор аттракторов несколько более разнообразен и определяется сигнатурой ляпуновских характеристических показателей так, как это показано в табл. 5.1.

5.4.6. Фрактальные свойства странного аттрактора. Рассмотрим подробнее характеристики странного аттрактора. Поскольку странный аттрактор не является подмногообразием, его геометрическая структура достаточно сложна. В частности, одной из его особенностей, часто встречающейся на практике, является масштабная инвариантность, состоящая в том, что при увеличении масштаба некоторой подобласти странного аттрактора мы получим геометрический объект, сходный по своей структуре с целым аттрактором. Физики называют такие объекты скейлинговыми структурами. Процедура масштабных преобразований может быть представлена как

[1] Напомним, что для симметричной матрицы сингулярный вектор совпадает с собственным.

разглядывание объекта через микроскоп с возрастающим увеличением; скейлинговые структуры приводят при этом к одинаковым изображениям при изменении увеличения. В этом смысле такие структуры называют самоподобными. С эффектами самоподобия мы встречались и при изучении логистического отображения и универсальности Фейгенбаума.

В математике геометрической моделью самоподобной структуры является фрактал, причем фракталами принято называть самоподобные структуры, обладающие дробной размерностью Хаусдорфа. К определению фрактала мы вернемся позже, здесь же только поясним, чем отличается обычная топологическая размерность от размерности Хаусдорфа. Считается, что топологическая размерность таких множеств, как прямая или ее отрезок, равна единице, размерность поверхности равна двум, объемного тела — трем и т.д. Однако есть более сложные геометрические объекты, для которых понятие размерности ввести затруднительно. Примерами таких объектов служат, например, траектория броуновской частицы, множество точек береговой линии, фазовая траектория движения с перемешиванием, странный аттрактор и др.

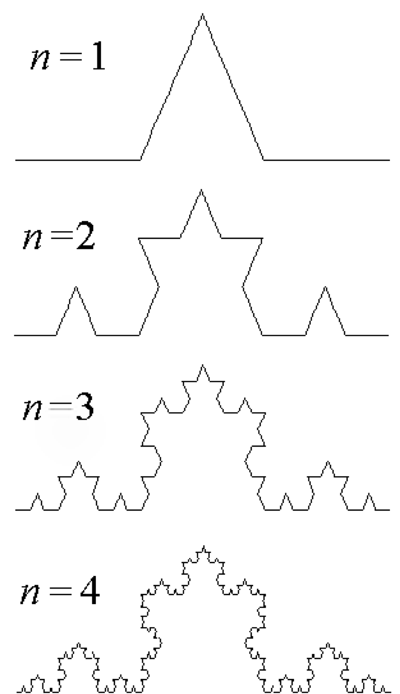


Рис. 5.23. Построение кривой Коха

Если бы была возможность найти топологическую характеристику, описывающую степень сложности такого множества точек, то можно было бы надеяться на то, что с ее помощью можно будет различать движение с перемешиванием от просто эргодического движения, охарактеризовать степень хаотизации движения и т.п.

Одной из таких возможностей является размерность Хаусдорфа. Пусть D — некоторое подмножество n-мерного пространства $\mathcal{R}_n$. Будем рассматривать покрытие множества D кубами из $\mathcal{R}_n$, плотно упаковывая их. Таких кубов возьмем столько, чтобы полностью покрыть D. Пусть r — сторона куба, и $N(r)$ — число кубов, требуемое для покрытия D (т.е. в это число попадают те и только те кубы, которые содержат хотя бы одну точку из D). Хаусдорфовой размерностью области D назовем

$$h_n = \lim_{r\to 0} \frac{\ln N(r)}{\ln(\frac{1}{r})}.$$

В качестве примера множества, имеющего дробную размерность, рассмотрим кривую Коха (рис. 5.23). Она получается в результате бесконечной процедуры, на каждом шаге которой на отрезке прямой выбирается его средняя треть, и заменяется на отрезок ломаной. Если длина начального отрезка прямой равна единице, то на n-м шаге длина

отрезочка станет равной $r_n = 1/3^n$. Для покрытия кривой Коха на n-м шаге достаточно использовать кубы длиной ребра, равной r_n, таких кубов будет 4^n. Таким образом,

$$h_n = \frac{\ln(4)}{\ln(3)} \approx 1.26\ldots$$

Позже, в главе, посвященной фракталам, мы вернемся к определению фрактальной размерности, не зависящей от формы ячейки покрытия.

5.5. Ведущие параметры

5.5.1. Быстрое и медленное время. До сих пор мы рассматривали в основном динамические системы с небольшим числом степеней свободы. Однако в реальных ситуациях нередко встречаются динамические системы, число степеней свободы которых достигает нескольких сотен и более. Как быть в таких ситуациях? Некоторые ответы на этот вопрос мы уже получили, когда рассматривали, например, метод усреднения, позволяющий следить за медленной эволюцией системы благодаря сглаживанию по времени. Динамику быстро изменяющихся координат можно исследовать методами сингулярно возмущенных уравнений.

В этом разделе мы коротко рассмотрим системы с большим числом степеней свободы и с более общих позиций покажем, как можно сократить это число выделением нескольких параметров, значения которых в основном и определяют эволюцию системы.

Динамику поведения сложных систем иногда можно представить как набор взаимосвязанных процессов, причем каждый процесс обладает своим собственным характерным ритмом, скоростью протекания. На структуре эволюционных уравнений это отражается таким образом, что их можно записать в виде системы уравнений вида

$$\mu^2 \frac{dx_i}{dt} = F_i(x_1, \ldots, x_n), \quad i = 1, \ldots, l, \tag{5.20}$$

$$\mu \frac{dx_j}{dt} = F_j(x_1, \ldots, x_n), \quad j = l+1, \ldots, l+m, \tag{5.21}$$

$$\frac{dx_k}{dt} = F_k(x_1, \ldots, x_n), \quad k = l+m+1, \ldots, n, \tag{5.22}$$

или, соответственно, в виде

$$\dot{x}_i = \frac{1}{T_1} F_i(x_1, \ldots, x_n), \quad i = 1, \ldots, l, \tag{5.23}$$

$$\dot{x}_j = \frac{1}{T_2} F_j(x_1, \ldots, x_n), \quad j = l+1, \ldots, l+m, \tag{5.24}$$

$$\dot{x}_k = \frac{1}{T_3} F_k(x_1, \ldots, x_n), \quad k = l+m+1, \ldots, n, \tag{5.25}$$

где $T_1 = \mu^2$, $T_2 = \mu$, $T_3 = 1$, коэффициенты μ^2 и μ определяют скорости протекания соответствующих процессов. Пусть $\mu \ll 1$. Если интересоваться поведением системы на малых интервалах времени длиной порядка $T_1 = \mu^2$, то переменные x_j, $j = l+1, \ldots, l+m$, и x_k, $k = l+m+1, \ldots, n$ за эти промежутки времени не успевают заметно измениться, и в уравнениях (5.20)–(5.22) и (5.23)–(5.25) их можно считать константами. Благодаря этому может существенно уменьшиться размерность исследуемой динамической системы. Если исследователь интересуется явлениями с характерной длительностью порядка $T_2 = \mu$, то в этой ситуации уравнение (5.20) и (5.23) описывает слишком быстрые, а (5.22) и (5.25) — слишком медленные процессы, поэтому в уравнениях (5.21) и (5.24) первые l переменные можно усреднить, а последние $n - l - m$ переменные можно считать константами. (В случае, когда среди первых l уравнений есть такие, которые определяют затухающие процессы, на временах порядка T_2 эти процессы стабилизируются, и соответствующие координаты можно будет заменить их стационарными значениями.) Аналогично поступаем, если нас интересуют характерные времена процессов порядка $T_3 = 1$. Заметим, что системы уравнений (1) и (2) являются сингулярно возмущенными, идеи решения таких задач нами рассматривались в разделе 3.2.2.

5.5.2. Параметры порядка и принцип подчинения. Системы с большим числом степеней свободы при изменении управляющих параметров могут проходить через регулярные и хаотические режимы поведения, причем периоды хаоса сменяются периодами относительного порядка, которые на макроскопическом уровне проявляются в возникновении некоторых макроструктур. Примером такого движения служит развитие турбулентности, на микроуровне демонстрирующее хаотическое движение молекул, переходящее в упорядоченное, на макроуровне проявляющееся в виде турбулентных вихрей.

Такое поведение системы наводит на мысль о существовании некоторого числа выделенных параметров, описывающих среду, которые, увеличиваясь со временем (или с изменением управляющих параметров), в основном и определяют состояние среды в каждый момент времени. Эта догадка послужила основой принципа подчинения, сформулированного немецким ученым Г.Хакеном, автором термина "синергетика", обозначающего дисциплину, изучающую принципы самоорганизации в сложных нелинейных открытых системах.

Проявление принципа подчинения мы видим и на рассмотрен-

ном выше примере уравнений (5.20)–(5.22) и (5.23)–(5.25): в нем на разных этапах эволюции динамической системы главную роль играют не все ее координаты, а лишь некоторое их подмножество. Оказывается, что в ряде случаев можно определить набор координат, через значения которых могут быть выражены значения остальных координат динамической системы. Тогда эти главные, определяющие координаты системы носят название параметров порядка. Говорят, что остальные координаты подчинены параметрам порядка. Выбор параметров порядка позволяет существенно снизить размерность исследуемой системы. Набор параметров порядка меняется с течением времени, причем смена параметров порядка происходит, как правило, резко, скачком. Рассмотрим еще один пример, поясняющий принцип подчинения и понятие параметров порядка.

Пусть динамика некоторой динамической системы описывается эволюционным уравнением

$$\dot{\mathbf{x}} = F(\mathbf{x}). \tag{5.26}$$

Рассмотрим систему, находящуюся вблизи седловой точки $\mathbf{x}_0$ фазового пространства $\mathcal{R}_N$, и пусть N-мерная матрица A линеаризации правой части симметрична, и следовательно, имеет полный набор из N собственных ортонормированных векторов $\{\mathbf{q}_i\}$, отвечающих собственным значениям $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_m \leq 0$; $0 < \lambda_{m+1} \leq \ldots \ldots \leq \lambda_N$. Обозначим собственные векторы, отвечающие неположительным собственным значениям A, индексом s: $\mathbf{q}_{s1}, \mathbf{q}_{s2}, \ldots, \mathbf{q}_{sm}$; а положительным — индексом u: $\mathbf{q}_{u(m+1)}, \ldots, \mathbf{q}_{uN}$. Разлагая вектор $\mathbf{x} - \mathbf{x}_0$ по базису $\{\mathbf{q}_i\}$, получим

$$\mathbf{x} - \mathbf{x}_0 = \sum_{i=1}^{m} \alpha_{si}\mathbf{q}_{si} + \sum_{i=m+1}^{N} \alpha_{ui}\mathbf{q}_{ui};$$

здесь индекс s обозначает устойчивое направление, а индекс u — неустойчивое, координаты $\alpha_{s1}, \alpha_{s2}, \ldots, \alpha_{sm}$ вектора $\mathbf{x} - \mathbf{x}_0$ в базисе $\{\mathbf{q}_i\}$ экспоненциально убывают, а $\alpha_{u(m+1)}, \ldots, \alpha_{uN}$ — экспоненциально растут. С течением времени в основном поведении системы будет определяться значениями параметров α_{ui}, $i = m+1, \ldots, N$, причем может оказаться так, что число этих параметров значительно меньше, чем N — полная размерность фазового пространства; на рис. 5.24 изображено поведение системы в окрестности седловой точки $\mathbf{x}_0$ и вдали от нее в том случае, когда число неустойчивых параметров равно единице, а соответствующая положительному собственному значению λ_n собственная функция — тригонометрическая: $\sin(2\pi/T)t$. Параметры α_{ui}, $i = m+1, \ldots, N$, определяющие движение системы, являются параметрами порядка.

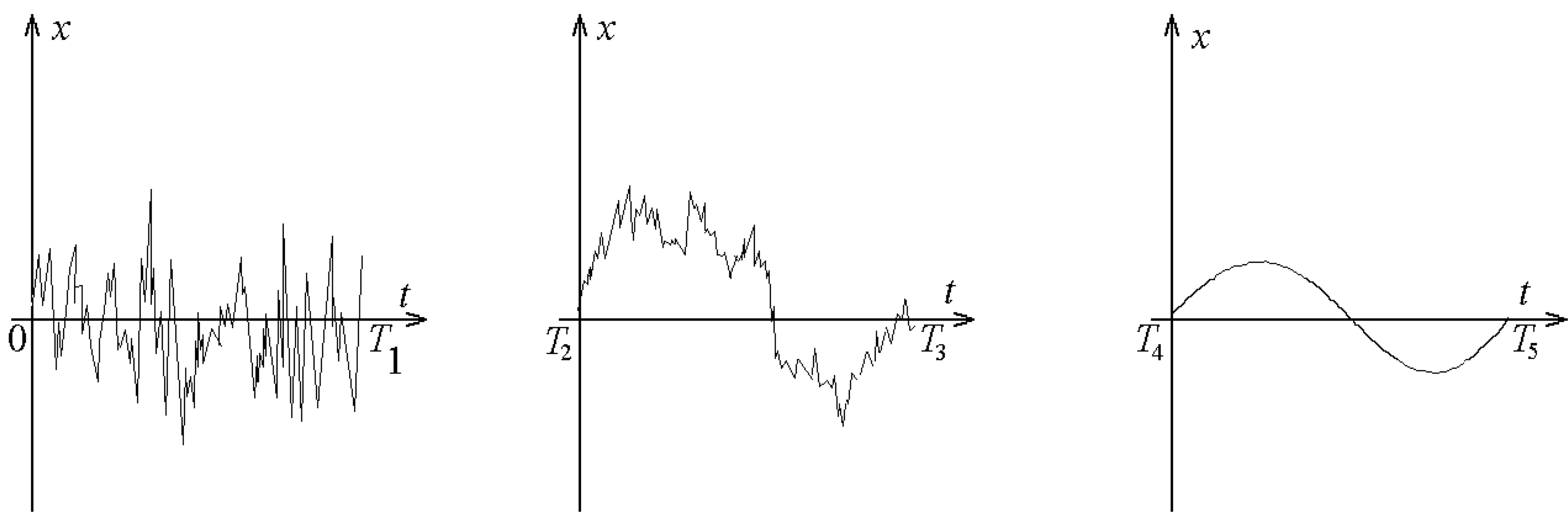


Рис. 5.24. Возникновение моды, описываемой ведущим параметром

С течением времени фазовая точка может попасть в окрестность другой седловой точки, где направление $\mathbf{q}_{un}$ может стать стабильным, тогда гармоническая мода, изображенная на рис. 5.24, затухнет. Если возникают другие неустойчивые направления $\mathbf{q}'_{ui}$, $i = m' + 1, \ldots, N$, являющиеся собственными векторами матрицы линеаризации A' правой части (5.26) в окрестности новой седловой точки, то параметры порядка меняются и возникают новые структуры.

Описанный здесь пример является сильным упрощением, связанным с рассмотрением линеаризованной системы. В общем же случае нелинейной системы не удается выделить независимые координаты, однако в ряде случаев устойчивые координаты как медленно меняющиеся в окрестности точки $\mathbf{x}_0$ можно выразить через неустойчивые; тем самым поведение системы будет полностью описываться лишь несколькими координатами α_{ui}, $i = m + 1, \ldots, N$ — параметрами порядка.

В качестве примера рассмотрим систему из двух обыкновенных дифференциальных уравнений:

$$\dot{u} = \alpha u - us, \tag{5.27}$$

$$\dot{s} = -\beta s + u^2, \tag{5.28}$$

где $\alpha \geq 0$, $\beta > 0$. Пренебрегая нелинейными членами, найдем, что u — неустойчивый параметр, а s — устойчивый. Уравнение (5.27) решается в квадратурах, так, что можно записать

$$s(t) = \int_{-\infty}^{t} e^{-\beta(t-\tau)} u^2(\tau) d\tau. \tag{5.29}$$

Здесь подразумевается, что $s(-\infty) = 0$. Если $u(t)$ при $t \to \infty$ растет медленнее экспоненты, то (5.29) имеет смысл. Однако такое подчинение неудобно, так как для $u(\cdot)$ получается интегро-дифференциальное уравнение (подстановкой (5.29) в (5.27)), и для вычисления $u(t)$ требуется знать все значения $u(\tau)$ при $\tau < t$. Однако можно привести

соотношение (5.29) к виду, в котором значения функций $s(\cdot)$ и $u(\cdot)$ связываются при одном и том же значении времени t. Для этого следует применить к (5.29) формулу интегрирования по частям, и получить разложение $s(\cdot)$ в ряд по степеням u [19].

Другой способ получения связи между s и u состоит в делении (5.28) на (5.27), в результате чего получается уравнение

$$\frac{\partial s}{\partial u} = \frac{u^2 - \beta s}{\alpha u - us},$$

которое можно рассматривать как обыкновенное дифференциальное уравнение первого порядка.

Принцип подчинения является эффективным средством исследования динамических систем с хаотическими режимами путем понижения размерности системы эволюционных уравнений.

Глава 6

Фракталы: определения и свойства

С понятием фрактала мы уже встречались в предыдущей главе, когда рассматривали фазовые траектории странного аттрактора. Однако эти объекты настолько интересны и с чисто математической, и с прикладной точек зрения, что заслуживают отдельного рассмотрения.

6.1. Что такое фрактал

6.1.1. Примеры и определения фрактала. Объекты, которые сейчас называются фракталами, впервые появились в математике при развитии понятия “линия”, “плоская фигура” и т.п.: к ним относятся такие фигуры, которые нельзя назвать в полном смысле слова ни линией, ни плоской фигурой. Примером такого объекта является уже рассмотренная нами в главе 5 кривая Коха, названная в честь датского математика Хельге фон Коха. Она получается из отрезка прямой последовательной заменой каждого прямолинейного участка на ломаную путем “вытягивания” средней трети исходного отрезка до равностороннего треугольника. Повторяя такую процедуру бесконечное число раз, в пределе получим конечную “линию”, соединяющую две точки, но имеющую бесконечную длину. Для привычных нам линий такое свойство кажется экзотичным. В то же время назвать кривую Коха плоской фигурой тоже не поворачивается язык — скорее, это “пушистая линия”.

Строго определить фрактал как математический объект не удается, есть лишь попытки дать такое определение. Наиболее известными являются определения Бенуа Мандельброта, математика, благодаря

работам которого теперь осознается, насколько важны эти новые геометрические объекты для понимания окружающего мира. В основе первого, пробного определения лежит представление о топологической размерности множеств: размерность точки принимается равной нулю, линии — единице, плоской фигуры — двум и т.п. Формулируется оно так: "Фракталом называется такое множество, размерность Хаусдорфа–Безиковича которого строго больше его топологической размерности". "Дробность размерности" и выражает "пограничное" свойство фракталов лежать между точкой и линией, или между линией и поверхностью и т.д.. Это как раз такие "пушистые" линии, как описанная выше кривая Коха. Однако, мало того, что требуется расшифровка понятия дробной размерности Хаусдорфа–Безиковича (к строгому определению этого понятия, обобщающего размерность Хаусдорфа, рассмотренную в разделе 5.4.6, мы вернемся позже), неудачность его стала очевидной после приведения ряда контрпримеров объектов, для которых это определение не выполняется, хотя их, исходя из интуитивного представления, имело бы смысл отнести к фракталам (например, чрезвычайно "дырявая" пирамида, построенная польским математиком, топологом Вацлавом Серпинским, формально имеет размерность, равную двум, хотя получена из трехмерного тетраэдра поочередным отбрасыванием вписанных в него тетраэдров с половинной стороной).

Несколько менее формальное и значительно более общее определение фрактала, данное Б. Мандельбротом несколько позже, звучит так: "Фракталом называется структура, состоящая из частей, которые в некотором смысле подобны целому". Неопределенность этого определения, содержащаяся в словах "в некотором смысле", делает понятие фрактала чуть ли не всеобъемлющим.

Поясним, как в это определение укладываются "математические" фракталы типа прямой фон Коха. Для этого вначале заметим, что такие геометрические объекты, как прямая или плоскость, разумно назвать самоподобными. Формально охарактеризуем это свойство можно тем, что эти фигуры не изменяются при некоторых геометрических преобразованиях: перенос прямой вдоль нее приводит к той же самой прямой, плоскость переходит в себя при параллельном сдвиге и повороте. Независимость от преобразований в математике называется симметрией. Есть множества, не обладающими столь полной симметрией, как плоскость или прямая, например, окружность не изменяется только при повороте — она тоже самоподобна. В этом смысле, согласно второму определению, все эти множества являются фракталами, несмотря на свою простую геометрическую структуру. Их можно назвать "гладкими фракталами", в отличие от кривой Коха, пирамиды Серпинского, множества Кантора и т.п. Какой же симметрией обладает кривая Коха? Выбрав ее фрагмент, например,

одну треть всей кривой, и увеличив его в три раза, мы вновь получим в точности исходную кривую. Физики говорят, что такие объекты обладают скейлингом, от слова scale — шкала: изменить шкалу в три раза — это все равно, что рассматривать исходный объект под микроскопом с троекратным увеличением. Если мы вновь видим ту же картину, что и без микроскопа, значит, объект обладает скейлингом и является фракталом.

6.1.2. Фрактальность пространственных форм. Сначала фракталы воспринимались как экзотика. Казалось, не бывает в природе объектов, для которых адекватной моделью является линия с бесконечной длиной, или объемная фигура с нулевым объемом. Но вот морское министерство Великобритании заказывает своему геодезическому управлению работу по измерению длины береговой линии Британских островов, и выясняется, что ответ зависит от масштаба используемой карты, имея тенденцию стремиться к бесконечности при уменьшении масштаба. Адекватной моделью береговой линии является фрактал. То же самое можно сказать и о рисунке речной сети на поверхности Земли, о структуре разломов в горных породах, о следах, оставляемых в диэлектрике высоковольтным разрядом при его пробое, о скоплении молекул, осаждаемых из раствора (они выглядят как длинные разветвленные мохнатые цепочки типа кораллов или снежинок), о замысловатых узорах, составленных из молекул одного вещества, расползающихся по поверхности другого — к ним относятся и ледяные рисунки на окнах, появляющиеся в морозные дни, — это все примеры природных фракталов.

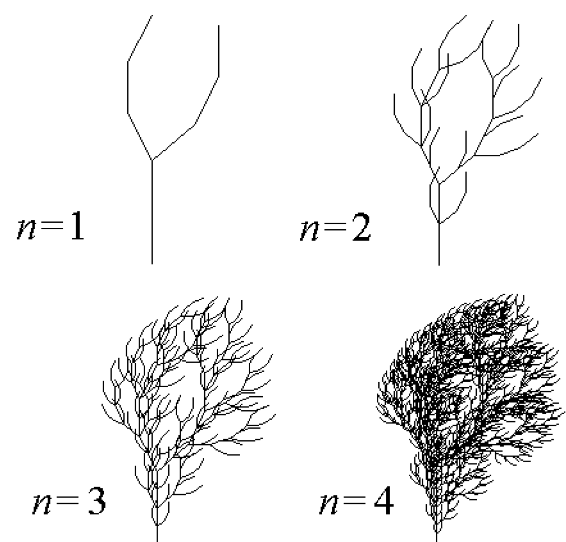


Рис. 6.1. Фрактальная структура куста

Фрактал обладает дробной размерностью. Во многих справочниках, особенно по материаловедению, часто можно встретить эмпирические зависимости типа степенной функции с дробным показателем, и для объяснения таких “странных” законов весьма правдоподобной кажется гипотеза, что эти зависимости отражают фрактальные свойства объектов, их порождающих — структуры зерен металла, структуры поверхностей и пр.

Кусты, деревья и другие растения тоже имеют фрактальную структуру. Действительно, рассмотрим процесс роста куста. Сначала появилась веточка, потом она выпустила два побега, на следующем этапе каждый побег вновь раздвоился, то же самое происходит на следующем этапе, и в результате из начальной “вилки” двух побегов вырастает причудливое растение, но самоподобное. Оно получено

многократным тиражированием простейших вилок, рис. 6.1.

Рассмотрим околоземное пространство — в нем есть центральное тело, Земля, вокруг которой вращаются спутники. Изменим масштаб, и получим ту же картину для Солнечной системы. Еще больший масштаб — и та же ситуация для галактики, для их скоплений и т.п. Выберем меньший масштаб — и получим сходную с этой картину структуры вещества. Вселенная — пространственный фрактал.

6.1.3. Динамические фракталы. До сих под шла речь о фрактальности пространственных форм. Однако самоподобие можно увидеть и в динамике процессов, протекающих во времени. Действительно, мы часто говорим о цикличности истории — государства, этносы, общественные структуры, целые цивилизации в своем развитии проходят через сходные этапы, "история повторяется".

В развитии Вселенной тоже бросается в глаза повторяемость этапов развития: сначала из единой точки рождается множество форм — фундаментальные частицы, ядра простейших элементов. Затем, после концентрации этой первоматерии, происходит следующий этап, в чем-то подобный первому: из простейших протонов и альфа-частиц рождается множество более тяжелых ядер. По тому же сценарию идет образование планет — от однородного, единого — к множеству разных форм: от атомов ("точек", размерность которых равна нулю) через линейные молекулы (единицы) к плоским (двойкам) и объемным структурам (тройкам). Так же — от "единиц", отдельных нервных клеток, через линейные цепочки — "двойки" к "тройкам", пространственным структурам, формируется нервная система в процессе эволюции живых организмов.

Фрактальность поведения сложных нелинейных систем сейчас также может рассматриваться как их неотъемлемое свойство. Если система достаточно сложна, то она в своем развитии обязательно проходит через чередующиеся этапы устойчивого и хаотического развития. Причем, как мы видели в главе 5, сценарии перехода от порядка к хаосу и обратно поддаются классификации, и вновь все многообразие природных процессов распадается на небольшое число качественно подобных. Один из таких сценариев, связанный с логистическим отображением, рассматривался в разделе 5.3.1, очень напоминает античные натурфилософские и мифологические сценарии рождения мира. Сначала, при некотором значении коэффициента пропорциональности, в системе имеется лишь одно устойчивое положение равновесия — Единое еще не начало свой путь творения. При изменении коэффициента r наступает момент, когда точка равновесия раздваивается, возникает два устойчивых состояния, в которых система пребывает по очереди, то в одном, то в другом, шаг за шагом по времени. Потом каждая из этих точек вновь раздваивается, и ситуация повторяется, сохраняя общий рисунок. Рано или поздно мно-

жество точек равновесия плотно заполняет все множество состояний, система переходит к хаосу, полностью разрушая свою структуру. Но затем, при дальнейшем росте параметра, из хаоса вновь возникает некоторое конечное число упорядоченных состояний, которые в конце концов "схлопываются" в единственное, и все начинается сначала. Напомним, что в математической модели этого явления обнаружено множество подобных, скейлинговых элементов; эти свойства подобия носят название универсальности Фейгенбаума.

Представления о похожести, фрактальности процессов развития закреплены и в мифах. Согласно древнегреческой мифологической традиции, мир рождается по этапам, в развитии которых видны сходные, подобные черты. Несколько поколений богов сменяют друг друга, на каждом этапе выстраивая свой космос, упорядоченную вселенную, по подобным принципам.

6.2. Размерность Хаусдорфа–Безиковича

Вернемся к математическому понятию фрактала и напомним его первое определение: "Фракталом называется такое множество, размерность Хаусдорфа–Безиковича которого строго больше его топологической размерности". Топологическая размерность точки по определению равна нулю, линии — единице, поверхности — двойке и т.п. Это понятие апеллирует к нашим представлениям о поверхности, линии, объемном теле и т.п., однако могут возникнуть затруднения — к какому типу отнести достаточно сложные геометрические объекты фрактального типа. Для определения размерности Хаусдорфа–Безиковича рассмотрим ряд примеров.

Зададимся вопросом: как измерить величину множества точек $\mathcal{L}$ метрического пространства? Разобьем пространство на ячейки с характерным размером (диаметром) δ (это могут быть шары, кубы и др.) и подсчитаем число ячеек, покрывающих множество. Уменьшая размер ячеек и следя за скоростью возрастания их числа, необходимого для покрытия множества $\mathcal{L}$, можно получить представление о размерности $\mathcal{L}$, в частности, вычислить "длину" множества, его "площадь", "объем" и т.п. Действительно, пусть $\mathcal{L}$ — спрямляемая кривая длины L_0. Выберем минимальное покрытие, т.е. такое, которое состоит из наименьшего числа ячеек (существование такого покрытия для компактного множества следует из леммы Гейне–Бореля). Число ячеек $N(\delta)$ в этом покрытии будет пропорционально отношению L_0/δ, и длину кривой получим предельным переходом при $\delta \to 0$:

$$L = N(\delta)\delta \sim \frac{L_0}{\delta}\delta \underset{\delta\to 0}{\longrightarrow} L_0.$$

Множеству точек, представляющих собой кривую длины L_0, можно сопоставить и площадь. Так как каждая ячейка имеет характерную площадь δ^2, то площадь множества пропорциональна $\delta^2 N(\delta)$, и при $\delta \to 0$ получим, что

$$S = N(\delta)\delta^2 \sim \frac{L_0}{\delta}\delta^2 = L_0\delta \underset{\delta\to 0}{\longrightarrow} 0.$$

Аналогично можно вычислить объем множества точек и т.д.

Если в качестве множества $\mathcal{L}$ взять часть квадрируемой поверхности площади S, то число ячеек, необходимое для покрытия этой поверхности, пропорционально $N(\delta) = S/\delta^2$; "длина" поверхности определяемая предельным переходом, равна

$$L = N(\delta)\delta \sim \frac{S}{\delta^2}\delta = \frac{S}{\delta} \underset{\delta\to 0}{\longrightarrow} \infty,$$

площадь S, $0 < S < \infty$, дается соотношением

$$S = N(\delta)\delta^2 \sim \frac{S}{\delta^2}\delta^2 = S,$$

а объем равен нулю:

$$V = N(\delta)\delta^3 \sim \frac{S}{\delta^2}\delta^3 = \frac{S}{\delta} \underset{\delta\to 0}{\longrightarrow} 0.$$

В этих примерах для характеристики "величины" множества точек $\mathcal{L}$ используется некоторая пробная функция $h(\delta) \sim \gamma(d)\delta^d$, которая определяет величину ячейки, — длину при $d = 1$, площадь при $d = 2$, объем при $d = 3$ и т.п., и "величина", или мера, множества $\mathcal{L}$ определяется как сумма "величин" всех ячеек, покрывающих $\mathcal{L}$:

$$M_d = \sum h(\delta).$$

Здесь константа $\gamma(d)$ зависит от формы ячейки и от того, какое свойство множества характеризует функция $h(\delta)$; например, если $h(\delta)$ — площадь круга, то $\gamma(d) = 2\pi$, а если площадь квадрата, то $\gamma(d) = 1$. При некотором d соответствующая мера M_d при $\delta \to 0$ равна либо 0, либо бесконечности, либо некоторому конечному положительному числу, и то значение d, при котором M_d не равна ни нулю, ни бесконечности, адекватно отражает топологическую размерность множества $\mathcal{L}$. Заметим, что результат не изменится, если в качестве значений d выбирать не только целые числа, но и любые действительные. Это соображение позволяет обобщить понятие

размерности, и дать следующее определение размерности, называемой размерностью Хаусдорфа–Безиковича:

Определение. Число d_{cr}, такое, что

$$\lim_{\delta\to 0} M_d = \begin{cases} 0, & d > d_{cr} \\ \infty, & d < d_{cr} \end{cases},$$

называется *размерностью Хаусдорфа–Безиковича.*

Здесь пока рассмотрены лишь достаточно простые множества, для которых размерность Хаусдорфа–Безиковича совпадает с топологической размерностью, однако существуют и более сложные геометрические объекты, для которых скачок меры M_d от нуля до бесконечности происходит при нецелых значениях d. К таким объектам относится, в частности, и рассмотренная ранее кривая Коха, изображенная на рис. 5.23.

Более наглядное представление о размерности Хаусдорфа–Безиковича можно получить, рассматривая $N(\delta)$ как функцию δ со степенной особенностью в нуле:

$$N(\delta) = \frac{\text{const}}{\delta^d} + \alpha(\delta), \tag{6.1}$$

где $\alpha(\delta)\delta^d \to 0$ при $\delta \to 0$. Размерность Хаусдорфа–Безиковича, как легко видеть, равна степени особенности d, в этом смысле размерность Хаусдорфа–Безиковича определяет скорость роста числа элементов минимального покрытия множества $\mathcal{L}$ при стремлении размера элемента к нулю. Из (6.1) получим, что при критическом значении d с точностью до бесконечно малых величин выполняется соотношение $N(\delta)\delta^d = \text{const}$, откуда

$$d_{cr} = \lim_{\delta\to 0} \frac{\ln N(\delta)}{\ln \frac{1}{\delta}}. \tag{6.2}$$

Рассмотренная ранее размерность Хаусдорфа обобщается в приведенном здесь определении. Обобщение касается того, что размерность Хаусдорфа–Безиковича не зависит от формы ячейки — будь то сфера, эллипсоид, куб или ячейка другой формы.

Вычислим размерности Хаусдорфа–Безиковича ряда множеств, которые являются типичными фракталами.

Размерность кривой Коха. Рассмотрим кривую фон Коха, изображенную на рис. 5.23.

Вообще говоря, определение фрактала для этой кривой будет выполняться лишь в пределе, когда длина отрезка прямой стремится к нулю, результат же, полученный на n-ом шаге, называется предфракталом; предфрактал представляет собой обычную спрямляемую линию.

При $\delta = 1/3$ число элементов покрытия равно $N(\delta) = 4$, а при $\delta = (1/3)^n$ — соответственно $N(\delta) = 4^n$, поэтому согласно (6.2) размерность Хаусдорфа–Безиковича кривой фон Коха равна

$$D = \frac{\ln 4}{\ln 3} > 1.$$

Размерность салфетки Серпинского. Вацлав Серпинский предложил пример "плоского" фрактального множества, получаемого из правильного треугольника последовательными выбрасыванием средних частей. Это множество изображено на рис. 6.2.

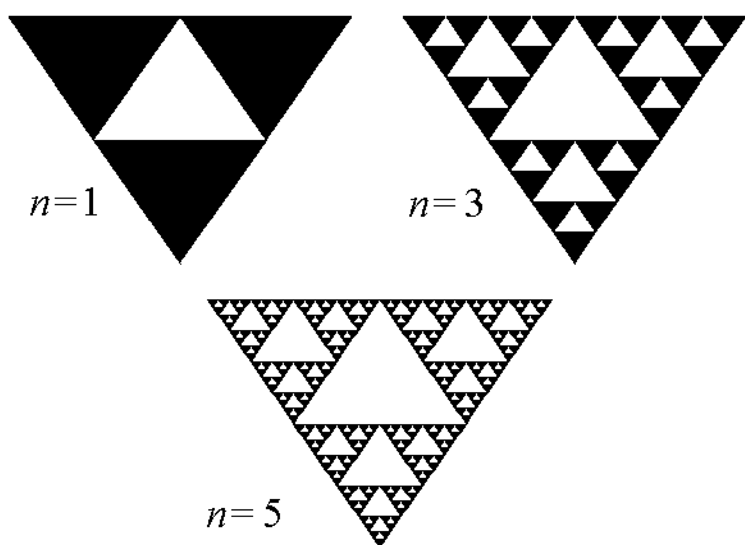


Рис. 6.2. Последовательное построение салфетки Серпинского

Размерность салфетки Серпинского легко подсчитать по формуле (6.2), выбирая в качестве элемента покрытия правильный треугольник со стороной $\delta = (1/2)^n$, тогда $N(\delta) = 3^n$, а значит, $D = \ln 3/\ln 2$. Заметим, что в этом случае $1 < D < 2$, и если считать "топологическую размерность" салфетки Серпинского равной двум (так как она получена из плоской фигуры), то первое определение фрактала не выполнено. С другой стороны, так как суммарная площадь всех выкинутых треугольников равна общей площади исходного треугольника, такое множество трудно назвать плоским, т.е. имеющим топологическую размерность, равную двум. Все это демонстрирует трудности, возникающие на пути применения первого определения фрактала.

Ковер Серпинского. Это тоже плоская фрактальная фигура, полученная способом, аналогичным способу получения салфетки Серпинского, но начальным элементом здесь является единичный квадрат, рис.6.3. При построении ковра Серпинского на первом шаге единичный квадрат делится на девять маленьких квадратов с длиной стороны, равной $1/3$, выбрасывается центральный квадрат и процедура повторяется с оставшимися квадратами бесконечное число раз. Размерность ковра Серпинского равна $D = \ln 8/\ln 3$, и $1 < D < 2$.

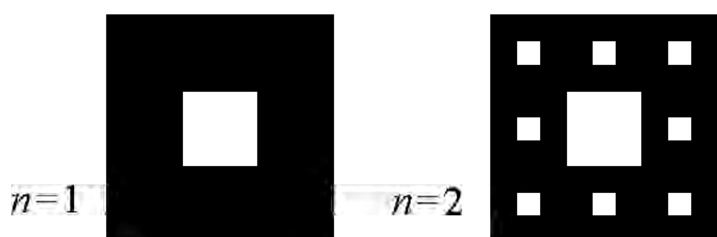


Рис. 6.3. Последовательное построение ковра Серпинского

Трехмерный аналог салфетки Серпинского. Определение фрактала как объекта с дробной размерностью Хаусдорфа–Безиковича тоже не выдерживает критики, что демонстрируется примером пирамиды

Серпинского. Это — объемная фрактальная фигура, полученная способом, аналогичным способу получения плоской салфетки Серпинского, но начальным элементом здесь является правильный тетраэдр с единичной длиной ребра. Фигура получается из тетраэдров, последовательно отсекаемых от вершин исходного тетраэдра, стороны отсекаемых тетраэдров равны половине стороны тех тетраэдров, от которых они отбрасываются.. При элементе покрытия в виде тетраэдра с длиной стороны $\delta = (1/2)^n$ требуется $N(\delta) = 4^n$ элементов покрытия, что приводит к размерности пирамидки Серпинского, равной двум. Итак, размерность этого объекта — целое число, хотя назвать его "плоским" вряд ли возможно.

6.3. Фрактал как самоподобный объект

Вернемся ко второму определению фрактала и рассмотрим подробнее свойства самоподобия.

Рассмотрим ограниченное множество, например, отрезок прямой. Он не обладает трансляционной симметрией, однако если изменить его длину в $r < 1$ раз, то параллельно перенося его несколько раз вдоль исходного отрезка прямой, им можно накрыть часть исходного множества. Если $r = 1/N$, где N — целое число, то за N параллельных переносов им можно целиком накрыть исходный отрезок.

Прямоугольный участок можно накрыть N его уменьшенными копиями, если коэффициент подобия $r = 1/\sqrt{N}$, прямоугольный параллелепипед — если $r = 1/\sqrt[3]{N}$, и т.д. В общем случае, если N уменьшенных копий исходного объекта с коэффициентом подобия $r = 1/N^{1/d}$ полностью накрывают его, то число d называют размерностью подобия, а сам объект, в силу приведенного выше определения, является фракталом. Из определения размерности подобия легко получить, что

$$d = \frac{\ln N}{\ln \frac{1}{r(N)}}. \tag{6.3}$$

В качестве примера вновь рассмотрим кривую Хельге фон Коха. С масштабным множителем $r = \frac{1}{3}$ мы получим фрагмент, состоящий из $\frac{1}{3}$ кривой, и таких фрагментов нужно 4, чтобы покрыть всю кривую. Аналогично, при $r = (\frac{1}{3})^n$ нужно 4^n фрагмента; следовательно, размерность подобия кривой Коха равна $\ln 4/\ln 3$ и совпадает с фрактальной размерностью Хаусдорфа–Безиковича.

Самоподобие и масштабирование в физических системах связано с существованием некоторых общих принципов, лежащих в основе

строения и развития этой системы. Использование симметрии существенно упрощает описание объекта, так как сводит различные проявления к единой причине. С симметрией и масштабированием (скейлингом) мы уже встречались, например, когда шла речь о логистическом отображении, см. раздел 5.3.1, которое демонстрирует скейлинговую структуру точек бифуркации, характерную для любого локально квадратичного отображения — речь идет об универсальности Фейгенбаума, выражающей принцип подобия. Анализ поведения квадратичных отображений приводит к одному из универсальных сценариев перехода к хаосу через удвоение периода.

6.3.1. Фракталы Жюлиа и Мандельброта. Двумерный аналог универсальности квадратичного отображения на плоскости исследован Б.Мандельбротом. Его идея состояла в том, чтобы вместо действительных чисел использовать комплексные и наблюдать процесс $x_1, x_2, \ldots$ не на прямой, а на плоскости.

Рассмотрим отображение

$$x_{n+1} = f(x_n) = x_n^2 + c, \tag{6.4}$$

где $x, c \in \mathcal{C}$ — комплексные числа, и рассмотрим множество аттракторов на комплексной плоскости. С изменением параметра c изменяются и области притяжения аттракторов.

Например, если $c = 0$, т.е. два аттрактора — ноль и бесконечность. "Бассейном" притяжения нуля являются точки, лежащие внутри единичной окружности, а областью притяжения бесконечности — точки, лежащие вне ее. Неустойчивые неподвижные точки, неустойчивые циклические точки отображения (6.4), а также точки, демонстрирующие хаотическое поведение последовательности $\{x_n\}$, также неустойчивые, и все их образы, лежат на окружности радиуса единица. Если же $c \neq 0$, то ситуация в чем-то похожая, так как по-прежнему могут существовать 2 или более аттракторов, один из которых — бесконечно удаленная точка, однако граница между областями притяжения уже не является гладкой и демонстрирует свойства фрактала. Одной из характерных черт этой границы является ее самоподобие: если изобразить ее в большем масштабе, то она будет иметь черты, схожие с полным изображением.

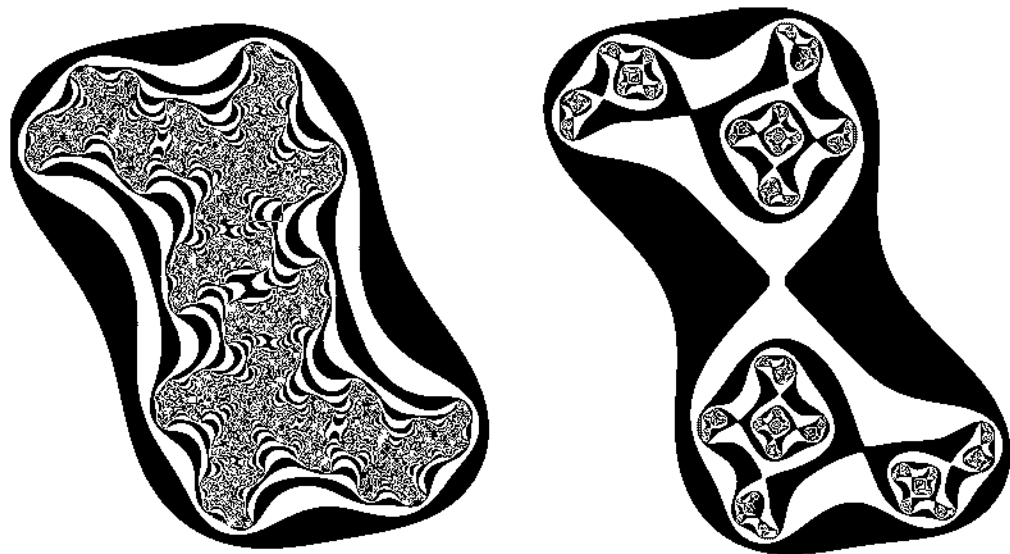


Рис. 6.4. Множества Жюлиа с различным числом бассейнов притяжения

Множества такого рода в математике известны как множества

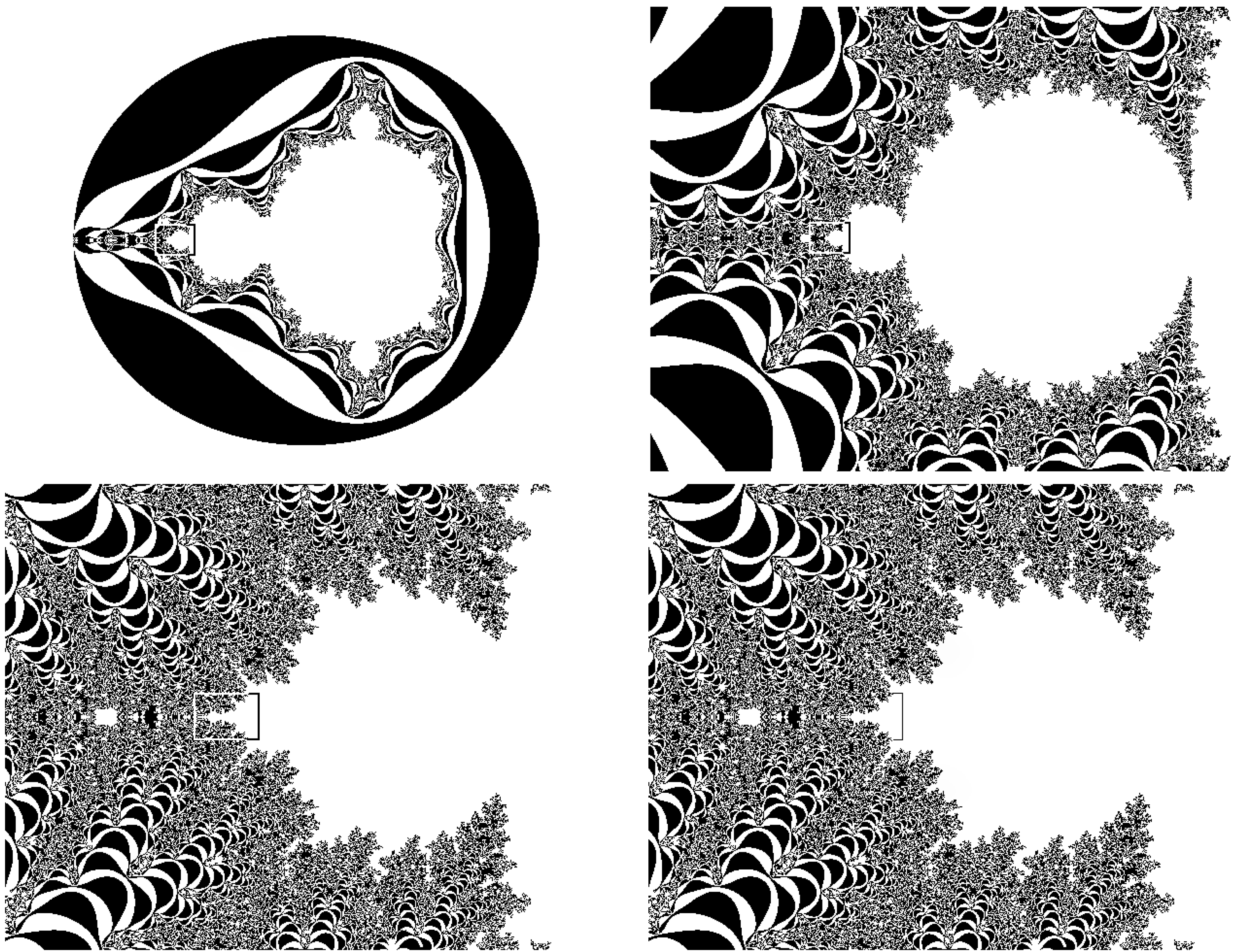


Рис. 6.5. Множество Мандельброта. Показаны увеличенные фрагменты предыдущего изображения, обведенные в рамку. Видна голографическая структура множества Мандельброта

Жюлиа, они изучались французскими математиками Пьером Фату и Гастоном Жюлиа еще в начале XX века. Множество Жюлиа для квадратичных отображений представляет собой множество всех неустойчивых неподвижных точек, неустойчивых циклических точек отображения (6.4), а также неустойчивых точек, демонстрирующих хаотическое поведение последовательности $\{x_n\}$, и всех их образов. Примеры множества Жюлиа для двух и большего числа аттракторов приведены на рис. 6.4, где линиями уровня отделены области, отличающиеся скоростью стремления последовательности $\{x_n\}$ к своему аттрактору.

В зависимости от параметра c может существовать несколько устойчивых неподвижных точек (аттракторов) отображения (6.4); и множество Жюлиа может иметь весьма экзотическое строение. Существует правило, показывающее, какой вид имеет множество Жюлиа при том или ином c, оно приводит на комплексной плоскости к множеству Мандельброта M, рис. 6.5, тоже имеющему фрактальные свойства. В зависимости от того, принадлежит c множеству M или нет, существуют различные множества Жюлиа. Если $c \in M$, то множество Жюлиа представляет собой связную структуру, а при

$c \notin M$ множество Жюлиа рассыпается на бесконечное число кусочков. Поэтому граница множества Мандельброта наиболее интересна — при переходе параметра c через нее множество Жюлиа как бы взрывается, превращаясь "в пыль". При изображении множества Мандельброта с увеличением масштаба можно заметить многократное повторение целого множества в более мелких деталях. На рис. 6.5 приведены несколько фрагментов множества Мандельброта в различных масштабах, наглядно демонстрирующие голографическую структуру множества M: часть целого несет в себе память о всем целом.

Симметрия, о которой идет речь в определении фрактала, не обязательно подразумевает геометрическое подобие. Во фрактале могут сохраняться некоторые другие характеристики объекта, такие, как углы поворота линии, или некоторые усредненные характеристики — вероятность поворота линии на определенные углы при изменении масштаба и др.

6.3.2. Фрактальные кластеры. В определении фрактальной размерности существенным является стремление к нулю диаметра ячейки покрытия, однако при использовании фракталов в физике следует иметь в виду, что составные части физической системы обладают некоторыми минимальными размерами R_0 — имеется в виду, например, радиус молекулы или атома и т.д. Поэтому на практике для характеризации физической системы используют понятие размерности кластера. Оно описывает асимптотику поведения числа ячеек $N(\delta)$ в покрытии кластера (совокупности объектов, составляющих физическую систему) в зависимости от их характерного размера, так же как и понятие размерности Хаусдорфа–Безиковича; отличие здесь состоит лишь в том, что зависимость $N(\delta)$ имеет реальный смысл при $\delta > R_0$, и размерность кластера характеризует лишь тенденцию скорости роста $N(\delta)$ при уменьшении δ.

Формально размерность кластера D может быть получена как показатель степени в формуле

$$N(\delta) \sim \rho \left(\frac{R}{\delta}\right)^D , \tag{6.5}$$

где $\rho = \mathrm{const}$, связывающей число ячеек $N(\delta)$, покрывающих кластер, с размером ячейки δ и радиусом R наименьшей сферы, содержащей кластер.

Строго говоря, природные фракталы, описанные в разделе 6.1.2, на самом деле являются примерами фрактальных кластеров.

6.3.3. Фракталы как модели физических систем.

Протекание жидкости в ячейке Хеле-Шоу. "Вязкие пальцы". Ячейка Хеле-Шоу состоит из двух прозрачных пластин, разделенных зазором. Хеле-Шоу исследовал протекание жидкости в такой ячейке,

вводя в поток краситель и наблюдая линии тока. Такие ячейки удобно использовать для наблюдения фрактальной структуры поверхности раздела двух жидкостей разной вязкости, возникающей, когда более вязкая жидкость вытесняется менее вязкой. Граница раздела при этом оказывается неустойчивой и наблюдается прорастание "пальцев" вытесняющей жидкости в вытесняемую. С развитием процесса граница раздела приобретает весьма экзотический вид и обладает дробной размерностью. Теория образования вязких пальцев важна, например, для моделирования процесса вытеснения нефти водой при ее добыче из пористой породы.

Перколяция. Этот процесс моделирует просачивание воды через пористую среду. Отметим случайным образом некоторые узлы прямоугольной сетки — они будут моделировать поры в среде, случайным образом связанные между собой. Соседние отмеченные узлы будем считать сообщающимися, так, что жидкость, впрыснутая в какую-либо пору, может проникать лишь в соседнюю, связанную с ней. Поры, или узлы, связанные с выбранным центром втекания жидкости, образуют кластер. Как показали натурные и вычислительные эксперименты, размерности таких кластеров дробные, определяются плотностью пор и удобно представлять множество пор, занятых жидкостью, как фрактал. Для практики очень важны исследования, состоящие в определении доли ячеек, составляющих поры, необходимой для того, чтобы жидкость, впрыснутая в одну из ячеек, протекла до другой, находящейся от нее на заданном расстоянии. В вычислительных экспериментах исследуются характеристики пористости и фрактальной размерности кластеров, обеспечивающие заданные режимы протекания.

Оказалось, что перколяционные кластеры самоподобны. Если разрешение, при котором этот кластер наблюдается, "не слишком велико", то детали оказываются смазанными, однако кластер в целом остается подобным себе при большом разрешении, т.е. сохраняется общая структура кластера, в частности, в нем присутствуют "дырки" всех размеров и т.п. Это самоподобие реализуется и в симметрии формул, связывающих концентрации пор при разных масштабах, а также в инвариантности размерности кластера (6.5) при различных линейных размерах R_0 элементарных ячеек.

Аттрактор Эно. Примером динамической системы, для описания которой возникает необходимость анализа геометрической структуры, обладающей свойствами самоподобия и дробной размерностью, является странный аттрактор Эно. Для того, чтобы получить наглядное представление о поведении этой динамической системы, рассмотрим сечение Пуанкаре. Для аттрактора Эно координаты точек на

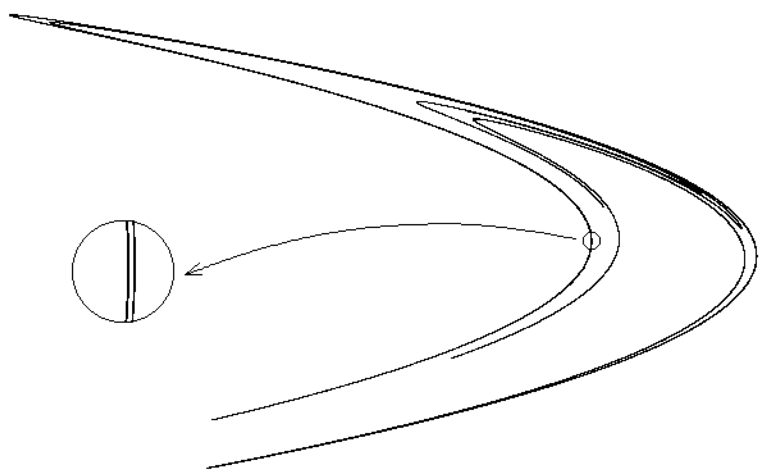


Рис. 6.6. Аттрактор Эно. Увеличенный фрагмент сохраняет структуру всего изображения

плоскости сечения Пуанкаре подчиняются системе уравнений

$$x_{n+1} = 1 - \alpha x_n^2 + y_n,$$

$$y_{n+1} = b x_n.$$

Множество точек пересечения фазовой траектории с поверхностью Пуанкаре для аттрактора Эно изображено на рис. 6.6. Точки этого множества сливаются в парные "линии", Однако при увеличении масштаба проявляются более тонкая внутренняя структура: можно заметить, что каждая линия из пары тоже состоит из двух "линий" В этом смысле можно говорить о фрактальной структуре отображения Пуанкаре для аттрактора Эно, так как тонкая структура сечения подобна его строению в целом.

6.4. Самоподобие как фундаментальное свойство природы

Мысль о том, что явления реального мира могут подчиняться математическим законам, возникла еще в античности. Язык математики тех времен был достаточно беден по сравнению с современным, его "словами" были числа и геометрические фигуры. Но уже тогда правила геометрии, применяемые при разметке участков земли или при строительстве, правила действия с числами при подсчете урожая, в астрономических расчетах или в торговле, давали точный ответ и никогда не подводили. Язык чисел и фигур был достаточно выразительным и универсальным, он позволял действительно находить общее, проявляющееся во многих явлениях реальности, на первый взгляд, казалось бы, совершенно не связанных между собой. Предсказательная сила, содержащаяся в математических моделях, настолько поражала ум (да поражает и сейчас, несмотря на привычку к современным техническим чудесам), что в древности за числам и математическими фигурами признавался мистический тайный смысл. Пифагор учил: "Что самое мудрое? — Число". Филолай из Кротона, его ученик, писал: "Все, что познаваемо, имеет число, без него ничего нельзя ни помыслить, ни познать". Платон в диалоге "Парменид, или Об идеях", а за ним и неоплатоники, в частности, Прокл, выстраивают иерархию Космоса от единого через "сверхсущие" единицы — непознаваемых богов (по сути — через числа) к "сущим", т.е. умопостигаемым богам.

Числа в древности несли в себе не только обозначение количества, но и великие принципы — Единство, Двойственность, Троичность и т.п., свойственные всему миру. Пользуясь числами как символами, античные философы описывали процесс рождения Космоса, т.е. то, как из Единого (обозначаемого единицей) возникает множественность форм.

Со временем мистический смысл математики теряется, на первый план выступает именно прикладной ее аспект. Но суть математики как всеобщего языка природы признается и поныне, мы надеемся, что пользуясь им, можно найти и выразить ту неуловимую общность, единое начало, исток всех явлений, которые связывают весь мир. На чем основана эта надежда? Еще в начальной школе мы узнаем, что число — это обозначение количества. Однако, числами можно характеризовать и качественные свойства мира — такие, например, как протяженность его объектов, тяжесть (вес) тел, высоту звука. Для этого требуется определить специальную процедуру измерения, задав единицу измерения, эталон, и определить способ подсчета количества эталонов, содержащихся в измеряемом объекте. Фундаментальное свойство природы — ее "измеримость", дает надежду на то, что на пути математической абстракции можно найти ответ на вопрос о том, в чем выражается общее, единое, что связывает разнородные явления мира.

Для того, чтобы числовая модель реальности не зависела от эталона, нужно рассматривать не абсолютные величины, а отношения количеств, пропорции. Но какие пропорции идеальны, а какие — нет? Вслед за античными мудрецами мы часто говорим о "божественной красоте" картины, или о "божественном звучании" музыки, не разделяя "божественное" и "прекрасное". Быть может, можно найти идеальные соотношения, опираясь на наше чувство красоты? По этому пути шли пифагорейцы, взяв за основу красоту созвучий — ведь отличить гармоничное звучание от какофонии может любой человек, не только музыкант. В пифагорейской теории музыки для анализа приятных на слух созвучий — консонансов — использовался инструмент, состоящий из одной струны, он назывался "монохорд". Наиболее гармоничное звучание получалось, когда звучали два монохорда, один с полностью открытой струной, другой — со струной, зажатой посредине. Это созвучие, называемое октавой, возникало, когда отношение длин звучащих струн (т.е. отношение высот двух звуков) равнялось двум. Два другие гармоничные созвучия получались при отношении длин струн два к трем (квинта) и три к четырем (кварта). И если чувство красоты человеку дано для ощущения "идеального", а законы прекрасного можно записать в виде математических соотношений, то появляется возможность находить соответствие "идеальному плану" как в явлениях природы, так и

в творениях рук человека: те объекты или явления, которые существуют по законам простых (целочисленных) пропорций, являются идеальными.

По дошедшей до нас легенде, свойства музыкальной гармонии настолько вдохновили Пифагора, что в отношениях целых чисел он стал искать главный ключ к законам мироздания. По его идее, весь мир пронизан вибрациями, и чтобы познать его, надо уметь услышать звуки мира, “музыку сфер”, прикоснуться к идеальной пропорциональности вселенских созвучий.

Еще одним ярким примером пропорции, закрепляющей чувство гармонии в строгих фиксированных математических законах, является так называемое “отношение золотого сечения”. Первое формальное ее определение содержится в “Началах” Евклида: “Говорят, что отрезок прямой разделен лучшим образом, пропорционально, если целая часть так относится к большей части, как большая к меньшей”. Отношение золотой пропорции встречается в природных объектах — в пропорциях человеческого тела, в строении раковины улитки, в сплетенной пауком паутине; в искусстве — архитектуре, живописи, скульптуре, музыке. Строение формы художественного произведения по законам золотой пропорции стало синонимом его совершенства. Парфенон в Афинах, храм Василия Блаженного в Москве, скульпторы античных мастеров — Фидия, Лисиппа, полотна Боттичелли, Рафаэля, Леонардо да Винчи, фуги Баха и сонаты Бетховена — везде присутствует золотое отношение.

Имея еще в древности столь блестящие подтверждения действенности математики в решении проблемы поиска единства явлений природы, человек продолжает искать новые объекты, новые законы, новые знаки и символы, отражающие общие принципы.

В рассмотренных примерах музыкальной гармонии и золотого сечения под пропорциями понимается отношение двух величин, измеренных с помощью одного и того же эталона. Равенство двух таких отношений выражает принцип подобия. Но подобие можно понимать и в более широком смысле. Например, все явления, описывающиеся дифференциальными уравнениями гиперболического типа, можно считать подобными, поскольку их поведение сходно на качественном уровне. Различные реализации случайного процесса тоже подобны, так как они описываются качественно одной и той же математической моделью. Можно считать, что современная наука только подтвердила, развила, наполнила новыми особенностями древний принцип: “все во всем”, или “что наверху, то и внизу”, который на современном языке можно понимать как указание на то, что мир самоподобен, т.е. что его части устроены так же, как и целое. Обозначением, символом самоподобия в современной математике и является относительно недавно возникшее геометрическое понятие — фрактал.

Глава 7

Численные методы исследования динамических систем

Заметим сразу, что если в динамической системе, эволюция которой описывается дифференциальными уравнениями

$$\dot{\mathbf{x}} = F(\mathbf{x}, \mathbf{c}) \tag{7.1}$$

с некоторым начальным условием

$$\mathbf{x}(0) = \mathbf{x}_0, \tag{7.2}$$

имеется неустойчивость, то исследовать ее численными методами путем прямого решения системы дифференциальных уравнений можно только на очень малых временах, где экспоненциальное разбегание изначально близких траекторий еще мало сказывается. Проверить этот факт можно, просчитав задачу "назад", то есть в качестве начальных условий задавая точку фазового пространства, полученную как значение приближенного решения в некоторый момент времени, а в эволюционных уравнениях обратить время, заменяя t на $-t$. Если обратный счет вернет фазовую траекторию в исходную точку (7.2), или в достаточно близкую к ней, то на этом интервале времени можно считать, что численное решение неплохо приближает точное. В противном случае численное решение будет отражать лишь дискретную схему, имеющую мало общего с исходной задачей (7.1), (7.2).

Как уже говорилось в разделе 2.2, при изучении поведения динамической системы часто интересуются не точными количественными характеристиками, а их качественным поведением. Поэтому численное исследование эволюционных уравнений не сводятся только к их численному интегрированию при различных начальных условиях и различных значениях параметра $\mathbf{c}$, а имеют смысл лишь в

сочетании аккуратными аналитическими методами исследования. В частности, часто используется численный поиск неподвижных точек системы (7.1) как решение нелинейного алгебраического уравнения

$$F(\mathbf{x}, \mathbf{c}) = 0$$

и численное исследование их зависимости от параметра **c**. Эта задача является основой для построения бифуркационных диаграмм. С помощью сочетания аналитических и численных методов находится отображение Пуанкаре, наглядно иллюстрирующее характер решения начальной задачи (7.1), (7.2). Численные методы линейной алгебры позволяют отвечать на вопрос об устойчивости решения и тем самым исследовать вопрос о структуре фазового пространства в зависимости от начальных условий и от параметров.

7.1. Расчет отображений Пуанкаре

Пусть автономная динамическая система задана эволюционным уравнением

$$\dot{x}_i = F_i(x_1, \ldots, x_N, \mathbf{c}), \quad i = 1, \ldots, N, \tag{7.3}$$

с начальными условиями

$$x_i(0) = x_{i0}, \quad i = 1, \ldots, N. \tag{7.4}$$

В N-мерном фазовом пространстве рассмотрим сечение Пуанкаре, задавая $(N-1)$-мерную поверхность S уравнением

$$S(x_1, \ldots, x_N) = 0. \tag{7.5}$$

Отображение Пуанкаре P преобразует точки поверхности S в точки той же поверхности, $\mathbf{x}_{n+1} = P(\mathbf{x}_n)$, где $\mathbf{x}_{n+1}$ и $\mathbf{x}_n$ — точки, лежащие на одной и той же фазовой траектории, рис. 7.1.

Для нахождения отображения Пуанкаре надо решать дифференциальное уравнение (7.3) с заданными начальными условиями и заданными значениями параметра **c**, и последовательно находить точки пересечения фазовой траектории с поверхностью S.

Здесь есть две различных задачи. Первая состоит в том, чтобы найти фазовую траекторию как решение начальной задачи (7.3), (7.4), а вторая — в том, чтобы определить координаты точки пересечения траектории с поверхностью S. Первая задача сама по себе достаточно сложна, если система (7.3), (7.4) неустойчива. Для ее решения рекомендуется использовать метод Рунге–Кутта четвертого

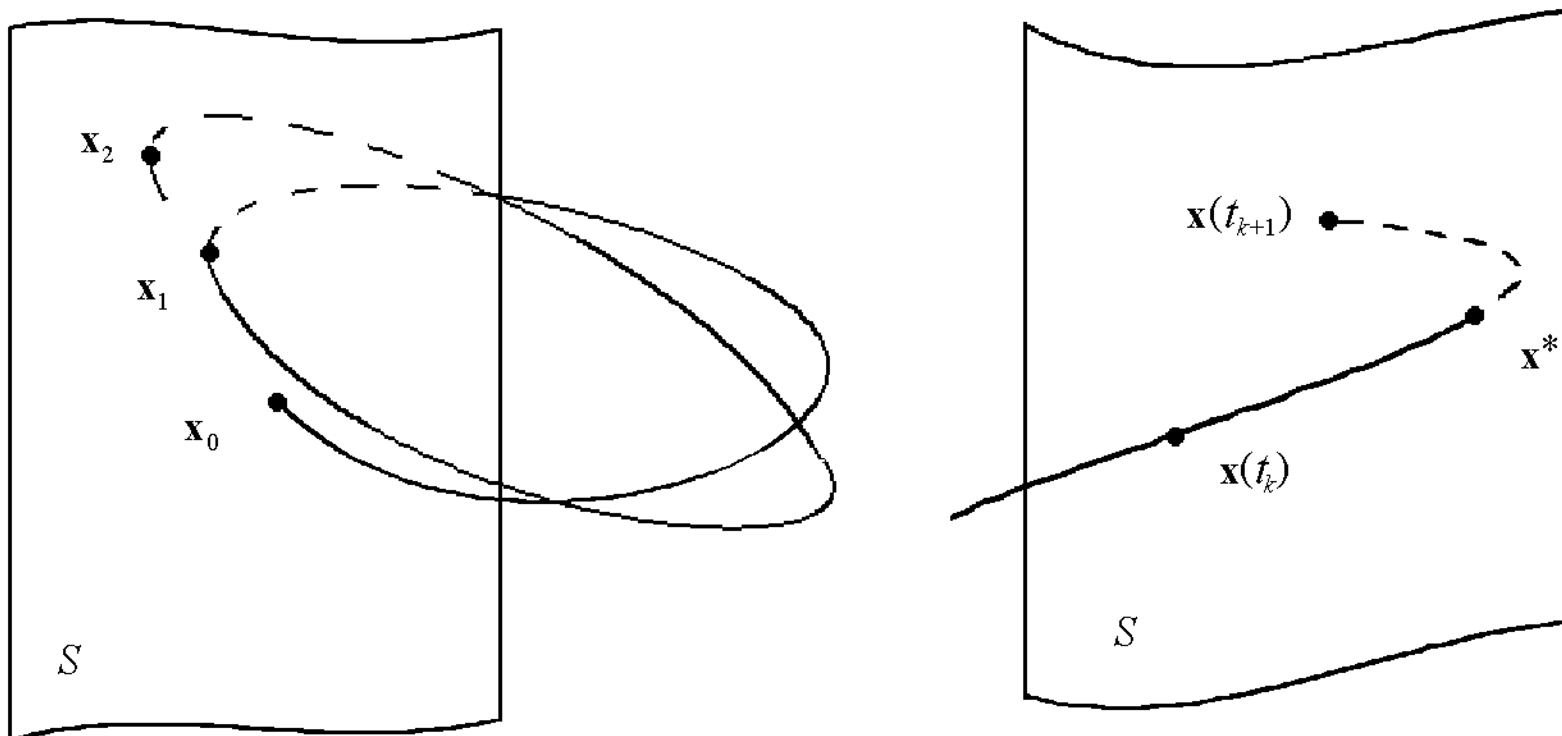


Рис. 7.1. Расчет отображения Пуанкаре

порядка и контролировать точность решения с помощью априорных и апостериорных оценок.

Более подробно рассмотрим метод определения координат точки пересечения фазовой кривой с поверхностью S. Для этого вычисляется знак функции S на каждом шаге интегрирования, и при смене его, например, с минуса на плюс мы имеем две точки траектории — $\mathbf{x}(t_k)$ и $\mathbf{x}(t_{k+1})$, лежащие по разные стороны от поверхности S. Точка пересечения траектории с S может быть найдена тем или иным способом интерполирования траектории между точками $\mathbf{x}(t_k)$ и $\mathbf{x}(t_{k+1})$. Рассмотрим здесь экономичный метод, позволяющий найти координаты точки пересечения путем интегрирования системы дифференциальных уравнений типа (7.3), (7.4) за один шаг.

Добавим формально к системе (7.3) еще одно уравнение относительно дополнительной переменной $x_{N+1} = S(x_1, \ldots, x_N)$:

$$\frac{dx_{N+1}}{dt} = \frac{dS}{dt} = \sum_{i=1}^{N} \frac{\partial S}{\partial x_i}\frac{dx_i}{dt} = \sum_{i=1}^{N} \frac{\partial S}{\partial x_i} F_i = F_{N+1}. \tag{7.6}$$

Теперь секущая поверхность Пуанкаре определяется уравнением

$$x_{N+1} = 0.$$

Поделим формально все уравнения (7.3) на уравнение (7.6), а последнее обратим, рассматривая t как функцию x_{N+1}, и получим

систему уравнений вида

$$\frac{dx_i}{dx_{N+1}} = \frac{F_i(\mathbf{x})}{F_{N+1}}, \quad i = 1, 2, \ldots, N,$$
$$\frac{dt}{dx_{N+1}} = \frac{1}{F_{N+1}}. \tag{7.7}$$

Теперь если $S(t_k) = S_1 < 0$, а $S(t_{k+1}) = S_2 > 0$, то координаты точки пересечения фазовой траектории с поверхностью S могут быть найдены из (7.7) на интервале изменения переменной x_{N+1} от S_2 до нуля, например, за один шаг длиной $-S_2$, с начальными условиями

$$x_i(x_{N+1})\,\big|_{x_{N+1}=S_2} = x_i(t_{k+1}), \quad i = 1, 2, \ldots, N;$$

$$t(x_{N+1})\,\big|_{x_{N+1}=S_2} = t_{k+1}.$$

В результате решения (7.7) на интервале изменения x_{N+1} длиной $|S_2|$ мы попадем в точку $x_{N+1} = 0$, т.е. на поверхность S.

7.2. Численный анализ периодических решений

Эти задачи состоят в нахождении периодических решений при фиксированном значении параметра **c** и определении их устойчивости, а также в изучении поведения семейства циклов в зависимости от параметра и построения бифуркационных диаграмм.

Для нахождения циклического решения надо иметь приблизительные данные о расположении циклической траектории в фазовом пространстве и о характере его устойчивости. Например, если цикл рождается из неустойчивого фокуса при бифуркации Андронова–Хопфа, то при малом отличии параметра от критического цикл следует искать вблизи неустойчивой особой точки.

В основу нахождения предельного цикла можно положить тот факт, что точка пересечения циклической фазовой траектории с некоторой поверхностью S есть неподвижная точка отображения Пуанкаре:

$$\mathbf{x}^* = P(\mathbf{x}^*). \tag{7.8}$$

Если задана точка $\mathbf{x}_0$ на поверхности S, находящаяся вблизи $\mathbf{x}^*$, то $\mathbf{x}^*$ можно найти с помощью итерационного алгоритма Ньютона как решение нелинейной системы уравнений (7.8). Причем оказывается,

что сходимость алгоритма Ньютона не зависит от характера устойчивости предельного цикла, а если этот цикл устойчив, то $\mathbf{x}^*$ можно найти методом простой итерации:

$$\mathbf{x}_{k+1} = P(\mathbf{x}_k). \tag{7.9}$$

Заметим, что искомый цикл может за один период пересекать поверхность S несколько раз (в разных точках); если $\mathbf{x}^*$ — такая точка, то это должно быть учтено в формуле (7.8).

Для исследовании устойчивости цикла определяют мультипликаторы цикла ρ_i, или, что то же самое, мультипликаторы неподвижной точки $\mathbf{x}^*$ отображения Пуанкаре. Для этого находят собственные числа оператора монодромии $Y(T)$, либо собственные числа линеаризованного отображения Пуанкаре в точке $\mathbf{x}^*$. Если мультипликаторы по модулю не превосходят единицу, то цикл устойчив по Ляпунову.

Для определения зависимости цикла от параметра уравнение (7.9) с начальным условием $\mathbf{x}^*(\mu)$ используется для нахождения предельной точки $\mathbf{x}^*(\mu + \Delta\mu)$, и все исследование повторяется сначала.

7.3. Вычисление спектра ляпуновских характеристических показателей

Полный спектр ляпуновских характеристических показателей несет достаточно подробную информацию о качественном поведении системы. Например, если в этом спектре, вычисленном в точке некоторого предельного цикла, имеются положительные показатели, то это — серьезное свидетельство в пользу того, что это предельное множество является странным аттрактором. Спектр ляпуновских характеристических показателей можно получить, исследуя сингулярные числа матрицы линеаризованного уравнения, описывающего эволюцию системы. Однако для качественного анализа динамики системы не столь важны абсолютные значения спектра ляпуновских характеристических показателей, сколько их знаки (сигнатура), а для их оценки можно предложить способ, не использующий численных методов линейной алгебры, основанный на следующих соображениях.

Если $\mathbf{x}_0 \in \mathcal{R}_N$ — неустойчивая точка системы (7.3), (7.4), то ее решение в окрестности $\mathbf{x}_0$ можно приближенно получить как решение линеаризованной задачи, представляющей собой сумму фундаментальных решений линеаризованной системы с коэффициентами, экспоненциально зависящими от времени (возможно, добавляются еще и степенные коэффициенты, не изменяющие экспоненциальную скорость роста решения). Анализируя решение линеаризованной

задачи при произвольно выбранном (случайном) начальном условии как зависящий от времени вектор N-мерного евклидова пространства, можно заметить, что через некоторое время основной вклад в него будет давать слагаемое с максимальной скоростью роста; это выражается в том, что через некоторое время направление вектора решения стабилизируется. Отклонения от этого направления при малых значениях времени t обусловлены заметной величиной и других слагаемых. Таким образом, можно приближенно оценить число линейно независимых составляющих решения, имеющих экспоненциальный рост. Для этого следует выбрать набор из N векторов из $\mathcal{R}_N$, представляющих собой последовательные решения линеаризованной задачи при различных значениях времени t, произвести их линеаризацию по процедуре Грамма–Шмидта, начиная с более поздних векторов, и исследовать ляпуновские характеристические показатели проекции решения на эти направления. Случайность начальных условий призвана обеспечить наличие в линеаризованном решении всех составляющих из набора фундаментальных решений. Аналогично можно поступать и с устойчивыми точками решений, только порядок выбора векторов при ортогонализации должен быть обратный. Подчеркнем, что точность вычисления спектра ляпуновских характеристических показателей в этом способе невелика, однако сигнатура оценивается достаточно точно.

Похожие соображения лежат в основе еще одного метода получения спектра ляпуновских характеристических показателей, приведенного в [1]. Прежде, чем перейти к его изложению, сделаем несколько замечаний.

Пусть $\mathcal{R}_N$ — евклидово пространство, $\mathbf{x} \in \mathcal{R}_N$, и задана функция

$$\mathbf{F}(t) = A(t)\mathbf{x} \equiv e^{\lambda t}\mathbf{x} \in \mathcal{R}_N, \tag{7.10}$$

где λ — некоторое число. Тогда для любого $\mathbf{y} \in \mathcal{R}_N$ ляпуновский характеристический показатель скалярного произведения $(\mathbf{F}(t), \mathbf{y})$ совпадает с показателем для функции $||\mathbf{F}(t)||$ для всех $\mathbf{y} \in \mathcal{R}_n$, за исключением тех, которые ортогональны $\mathbf{x}$, т.е. когда

$$(\mathbf{x}, \mathbf{y}) = 0. \tag{7.11}$$

Действительно,

$$\frac{1}{t}\ln||\mathbf{x}||e^{\lambda t} = \frac{1}{t}(\ln||\mathbf{x}|| + \lambda t) \to \lambda \quad \text{при} \quad t \to \infty$$

и

$$\frac{1}{t}\ln(\mathbf{x}, \mathbf{y})e^{\lambda t} = \frac{1}{t}(\ln(\mathbf{x}, \mathbf{y}) + \lambda t) \to \lambda \quad \text{при} \quad t \to \infty.$$

При случайном выборе $\mathbf{y}$ выполнение условия (7.11) маловероятно.

Второе замечание состоит в следующем. Пусть задана функция

$$\mathbf{F}(t) = \sum_{k=1}^{n} e^{\lambda_k t}\mathbf{x}_k \in \mathcal{R}_N, \tag{7.12}$$

где $\mathbf{x}_k \in \mathcal{R}_N$, а λ_k, $k = 1, \ldots, n$, — некоторые числа, упорядоченные по невозрастанию. Тогда для любого $\mathbf{y} \in \mathcal{R}_N$ ляпуновский характеристический показатель скалярного произведения $(\mathbf{F}(t), \mathbf{y})$ равен λ_1 — максимальному из всех чисел λ_k, $k = 1, \ldots, n$, и совпадает с показателем для функции $\|\mathbf{F}(t)\|$ для всех $\mathbf{y} \in \mathcal{R}_N$, за исключением тех, которые ортогональны $\mathbf{x}_1$. Это следует из соотношений

$$\frac{1}{t}\|F(t)\| = \frac{1}{t}\ln\left\{\sum_{\substack{k=1\\j=1}}^{n}(\mathbf{x}_k, \mathbf{x}_j)e^{(\lambda_k+\lambda_j)t}\right\}^{1/2} =$$

$$= \frac{1}{t}\ln\left\{e^{2\lambda_1 t}\sum_{\substack{k=1\\j=1}}^{n}(\mathbf{x}_k, \mathbf{x}_j)e^{(\lambda_k+\lambda_j-2\lambda_1)t}\right\}^{1/2} =$$

$$= \lambda_1 + \frac{1}{2t}\ln\left\{\sum_{\substack{k=1\\j=1}}^{n}(\mathbf{x}_k, \mathbf{x}_j)e^{(\lambda_k+\lambda_j-2\lambda_1)t}\right\} \underset{t\to\infty}{\longrightarrow} \lambda_1, \tag{7.13}$$

так как все показатели экспонент в коэффициентах суммы в формуле (7.13) неположительны, и

$$\frac{1}{t}\ln\left\{\sum_{k=1}^{n}(\mathbf{x}_k, \mathbf{y})e^{\lambda_k t}\right\} = \frac{1}{t}\ln\left\{e^{\lambda_1 t}\sum_{k=1}^{n}(\mathbf{x}_k, \mathbf{y})e^{(\lambda_k-\lambda_1)t}\right\} \underset{t\to\infty}{\longrightarrow} \lambda_1.$$

Наконец, рассмотрим прямое (внешнее) произведение векторов $\mathbf{y} \in \mathcal{R}_n$, $\mathbf{z} \in \mathcal{R}_m$: $\mathbf{y} \wedge \mathbf{z} \in \mathcal{R}_n \otimes \mathcal{R}_m$. Если

$$\mathbf{y} = \sum_{k=1}^{N}\mathbf{y}_k e^{\lambda_k t}, \quad \mathbf{z} = \sum_{k=1}^{N}\mathbf{z}_k e^{\mu_k t},$$

и $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$, $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_N$, то

$$\|\mathbf{y} \wedge \mathbf{z}\|^2 = \sum_{k,j=1}^{N}\mathbf{y}_j \wedge \mathbf{z}_k e^{(\lambda_j+\mu_k)t},$$

и скорость экспоненциального роста функции $\|\mathbf{y} \wedge \mathbf{z}\|^2$ равна $\lambda_1 + \mu_1$.

Эти замечания теперь используем для того чтобы, зная решение линеаризованной начальной задачи (7.3), (7.4), оценить суммы ляпуновских характеристических показателей ее фундаментальной системы решений, а затем — и сам спектр ляпуновских характеристических показателей. Для этого введем понятие обобщенного k-мерного ляпуновского показателя динамической системы (7.1), (7.2), [1].

Определение. *Обобщенным k-мерным ляпуновским показателем*, или показателем k-го порядка, будем называть число

$$\lambda^k_{\mathbf{x}_0} = \lim_{t\to\infty} t^{-1} \ln \left\| \bigwedge_{i=1}^{k} Y^t_{\mathbf{x}_0} \mathbf{e}^0_i \right\|, \tag{7.14}$$

где $\mathbf{e}^0_i$ — ортонормированный базис $\mathcal{R}_N$, $\wedge$ — внешнее произведение векторов, $\|\cdot\|$ — евклидова норма, $Y^t_{\mathbf{x}_0}$ — фундаментальная матрица решений линеаризованной задачи (7.1), (7.2).

В формуле (7.14) вектор

$$\mathbf{y}^t_{\mathbf{x}_0 i} = Y^t_{\mathbf{x}_0} \mathbf{e}^0_i$$

есть решение линейной задачи

$$\dot{\mathbf{y}}^t_{\mathbf{x}_0 i} = A(\mathbf{x})\mathbf{y}_{\mathbf{x}_0 i}, \quad \mathbf{y}_{\mathbf{x}_0 i}(0) = \mathbf{e}^0_i,$$

в момент времени t, $A(\mathbf{x})$ — матрица линеаризации системы (7.1).

Опишем свойства k-мерного ляпуновского показателя.

1. Число $\lambda^k_{\mathbf{x}_0}$ при фиксированном k может принимать не более, чем C^k_N различных значений.

2. При произвольном выборе базисных векторов $\mathbf{e}^0_i$ наиболее вероятно, что $\lambda^k_{\mathbf{x}_0}$ равно наибольшему из всех C^k_N своих возможных значений.

Действительно, если кратность показателя равна единице, то $\lambda^1_{\mathbf{x}_0}$ равен наибольшей реальной части собственного числа матрицы линеаризованной системы (7.1) (т.е. старшему ляпуновскому характеристическому показателю), если $\mathbf{e}^0_1$ не ортогонален соответствующему собственному вектору; если же эта ортогональность имеет место (что маловероятно при произвольном случайном выборе $\mathbf{e}^0_1$), но $\mathbf{e}^0_1$ не ортогонален следующему собственному вектору, то $\lambda^1_{\mathbf{x}_0}$ равен наибольшей реальной части собственного числа, следующей за первой, и т.д., всего возможно $N = C^1_N$ вариантов, все менее и менее вероятных. Если кратность равна двум, то в наиболее вероятном случае $\lambda^2_{\mathbf{x}_0}$ равна сумме двух наибольших ляпуновских характеристических показателей, и т.д. Таким образом, если не принять специальных мер, то $\lambda^k_{\mathbf{x}_0} = k\lambda_{max}$. Специальные меры в данном случае состоят в том, что в качестве векторов $\mathbf{e}^0_i$, выбираются ортогонализованные процедурой Грамма–Шмидта векторы $Y^t_{\mathbf{x}_0}\mathbf{e}^0_i$, причем этот выбор

время от времени обновляется. Тем самым обеспечивается нужная ортогональность векторов $Y^t_{\mathbf{x}_0}$: практически, векторы упорядочены по убыванию скоростей экспоненциального роста.

Действительно, дело в том, что при произвольном базисе $\{\mathbf{e}^0_i\}$ вектор $Y^t_{\mathbf{x}_0}\mathbf{e}^0_1$ может быть представлен в виде суммы

$$Y^t_{\mathbf{x}_0}\mathbf{e}^0_1 = \sum_{i=1}^{N} \mathbf{g}_i^{(0)} e^{\lambda_i t}, \quad \mathbf{g}_i^{(0)} \in \mathcal{R}_N, \tag{7.15}$$

и при $y \to \infty$ определяющую роль в сумме (7.15) будет играть слагаемое с максимальной скоростью экспоненциального роста. Ортогональные к нему слагаемые имеют временную зависимость вида $e^{\lambda_2 t}$, и т.д.

Изменение базиса не влияет на величину $\lambda^k_{\mathbf{x}_0}$, так как знаменатель в (7.14) для любого ортонормированного набора равен единице, и числитель, благодаря свойствам внешнего произведения векторов, также не зависит от выбора базиса $\mathbf{e}^0_i$.

Таким образом, зная наиболее вероятные значения всех k-мерных ляпуновских показателей, мы можем найти спектр характеристических показателей $\lambda_{\mathbf{x}_0,1}, \lambda_{\mathbf{x}_0,2}, \ldots, \lambda_{\mathbf{x}_0,N}$ из соотношений:

$$\lambda_{\mathbf{x}_0,1} = \lambda^1_{\mathbf{x}_0}, \quad \lambda_{\mathbf{x}_0,i} = \lambda^i_{\mathbf{x}_0} - \lambda^{i-1}_{\mathbf{x}_0}, \quad i = 2, 3, \ldots, N.$$

7.4. Расчет размерности аттрактора

7.4.1. Оценка топологической размерности. На практике часто возникает проблема анализа временной последовательности и исследования свойств динамической системы, основанных на наблюдениях за неполным набором координат фазового вектора, описывающего систему. Такие задачи встречаются в биологии и медицине, например, при анализе данных электроэнцефалограмм (ЭЭГ), при описании регенерации эритроцитов и т.п. Одной из важных характеристик динамической системы является размерность минимального пространства, содержащего аттрактор; эта размерность дает эффективное значение числа параметров, необходимых для достаточно полного описания поведения системы.

Для определения размерности аттрактора можно было бы использовать данные о фазовой траектории, которые содержатся, например, в последовательности наблюдений $\mathbf{x}(t_1), \ldots, \mathbf{x}(t_N)$ и воспользоваться процедурой вычисления фрактальной размерности кластера, однако

часто в экспериментальных исследованиях неизвестно описание процесса, не существует его модели и есть лишь ряд наблюдений за одним или несколькими параметрами, характеризующими систему. Однако оказалось, что при некоторых достаточно общих предположениях о характере поведения динамической системы можно оценить размерность аттрактора и из временного ряда наблюдений всего за одним параметром системы. Этот результат получен Такенсом в 1981 г. Рассмотрим лишь идею метода. Предположим, что динамическая система описывается эволюционным уравнением

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}), \quad \mathbf{x}(t) = (x(t), y(t)), \tag{7.16}$$

Значение вектора $\mathbf{x}(t+\tau) = (x(t+\tau), y(t+\tau))$ в момент времени $t+\tau$ полностью определяется значением вектора $\mathbf{x}(t) = (x(t), y(t))$ в момент времени t и уравнением (7.16), причем эта связь взаимно однозначна, если фазовая траектория не имеет самопересечений. Рассмотрим последовательность векторов

$$\begin{gathered} \xi(t) = [x(t), x(t+\tau)], \\ \xi(t+\tau) = [x(t+\tau), x(t+2\tau)], \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \xi(t+N\tau) = [x(t+N\tau), x(t+(N+1)\tau)]. \end{gathered} \tag{7.17}$$

Так как координаты вектора ξ связаны с координатами вектора $\mathbf{x}$ соотношениями

$$\xi_1(t) = x(t),$$

$$\xi_2(t) = x(t+\tau) = x(t) + \int_t^\tau F_1(x(t'), y(t'))dt' \sim x(t) + \tau F_1(x(t), y(t))$$

с якобианом, отличным от нуля, то информация о размерности аттрактора, содержащаяся в последовательности $\{\mathbf{x}(t_k),\ k = 1, \ldots, N\}$, присутствует и в последовательности $\{\xi(t + (k-1)\tau),\ k = 1, \ldots, N\}$. Опираясь на эти соображения, Такенс доказал, что для динамической системы, описывающейся моделью

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}),$$

где размерность вектора $\mathbf{x}$ равна M, метрические свойства M-мерного множества $\mathbf{x}(t_k)$, $k = 1, \ldots, N$, вложенного в пространство $\mathcal{R}_M$, эквивалентны метрическим свойствам множества $\{\xi(t + (k-1)\tau),$ $k = 1, \ldots, N\}$, где ξ — M-мерный вектор, определенный аналогично (7.17), точнее, расстояния между векторами того и другого множества связаны коэффициентами, отличными от нуля и равномерно ограниченными. Это соображение позволяет оценить характерную размерность аттрактора по характерной размерности множества

$\{\xi(t+(k-1)\tau),\ k=1,\ldots,N\}$. На практике эти свойства проявляются в свойствах автокорреляционной функции ряда наблюдений $x_j(t_k)$, $k=1,\ldots$

В качестве примера применения этой теоремы приведем результаты исследований, которые показали, что характерная размерность ЭЭГ человека в нормальном состоянии принимает значения между двумя и тремя, в то время как при эпилептическом припадке она резко падает и принимает значения, примерно равные единице.

7.4.2. Фрактальная размерность аттрактора. Для вычисления размерности странного аттрактора можно напрямую воспользоваться формулой для фрактальной размерности кластера, обсуждаемой в главе 6, однако более удобно для численных расчетов пользоваться формулой

$$D = N - \lim_{\delta\to 0}\frac{\ln V(\delta)}{\ln\frac{1}{\delta}},$$

где N — размерность фазового пространства, $V(\delta)$ — объем области фазового пространства, состоящей из всех точек, находящихся в δ-окрестности аттрактора.

Глава 8

Самоорганизация в нелинейных системах

8.1. О двух тенденциях динамики — от беспорядка к порядку и в обратном направлении

До сих пор мы говорили либо о вполне конкретных не слишком сложных физических системах, либо об абстрактных математических моделях. Но динамической, бразвивающейся системой является и окружающий нас реальный мир. Можно ли говорить о мире, о Космосе как о динамической системе и записать ее математическую модель? По-видимому, всеобъемлюще этот вопрос решить не удастся ни сейчас, ни в обозримом будущем, но можно попытаться ответить на вопросы об основных тенденциях движения мира.

Всем нам знакома ситуация, когда с любовью наведенный порядок в нашей квартире через несколько дней сменяется хаосом — это очень наглядная иллюстрация ко второму началу термодинамики, гласящему, что в замкнутой системе энтропия не убывает. Этот закон, открытый в XIX веке, наводит на мысль, что время в нашем мире всегда течет “в одну сторону”, это направление мы называем “от прошлого к будущему”. С другой стороны, есть и другие примеры, когда беспорядочное поведение системы вдруг обретает порядок — из хаотического теплового движения молекул возникают турбулентные вихри, из лишенных структуры скоплений достаточно простого по своему составу межзвездного вещества (из водорода и гелия) рождаются звезды, производящие в своих недрах более сложные химические элементы, рождаются и развиваются живые существа и т.п. Это примеры сложных нелинейных незамкнутых систем, которым

свойственна тенденция, противоположная предписываемой вторым началом. Что заставляет мир усложняться, несмотря на противодействие принципов равновесной термодинамики? Для объяснения этого феномена с древних времен привлекаются концепции творения, закрепленные в виде религиозных догм, однако наука идет путем разума и опыта, а не слепой веры, и на этом пути пытается найти ответы на вопросы, которые долгое время относились к философским или религиозным, открывая механизмы тех или иных явлений на все более и более глубоком уровне. Так, например, в свое время магнетизм, считавшийся тайной, открытой посвященным, предметом изучения оккультных наук, благодаря опытам Фарадея и теоретическим работам Максвелла потерял свою мистическую окраску, хотя вопрос о том, что такое электромагнитное поле и почему оно существует, остался не решенным (может быть, потому, что он сейчас не столь важен с практической точки зрения). В любой ситуации остаются проблемы причин возникновения тех или иных механизмов развития и существования Природы, но с развитием науки они отодвигаются на все более абстрактный план.

Современная теория нелинейных динамических систем также дает надежду на то, что будут открыты механизмы, приводящие от беспорядка к порядку, лежащие в основе творчества Природы. К настоящему моменту достигнуто определенное понимание роли неустойчивости и необратимости в самоорганизации — недавно осознанного свойства и живой, и неживой материи, состоящего в том, что при определенных условиях в системе, состоящей из хаотично расположенных неупорядоченных частей возникает новая структура, обладающая определенным порядком. Примером самоорганизации служат рассмотренные выше диссипативные структуры, в частности, возникающие в реакции Белоусова–Жаботинского, или вихри в ячейках Бенара. Эти примеры могут служить основой понимания и описания таких процессов, как возникновение жизни, развитие общества, кризисы в экономике и др.

Понятие самоорганизации исторически возникло достаточно давно, еще в религиозных и философских концепциях древности упорядоченный мир возникает как организация хаоса, толчком которому служит божественный импульс. После открытия примеров и законов самоорганизации самоупорядочение сложных систем можно рассматривать как фундаментальное свойство природы.

В естественных науках концепция самоорганизации появилась в середине XX века в кибернетике; где она первоначально понималась как процесс структурирования системы, управляемый изнутри самой системы. В этом проявлялся классический кибернетический подход, неклассический же подход, появившийся несколько позднее, допускал отсутствие блока управления, он возник в связи с рассмотрением

систем, в которых структурирование происходит самопроизвольно, спонтанно. Авторами такого подхода являются Н.Винер и Г. фон Фёрстер.

В 60–70 годах понятие самоорганизации как образования диссипативных структур выдвинул И. Пригожин. Такие структуры возникают в системах, находящихся вдали от термодинамического равновесия, условием их существования служат постоянный приток энергии извне и диссипация ее внутри системы. Заметим, что хотя именно в таких состояниях происходит существование живых организмов, появление диссипативных структур проявляется как общее свойство нелинейных систем с диссипацией.

В этом же направлении работал и Г. Хакен, основоположник синергетики, в которой тоже возникают явления самоорганизации в сильно неравновесных системах. Основным содержанием концепции Хакена является возможность описания систем в терминах параметров порядка, см. раздел 5.5.2.

Концепция самоорганизации тесно связана с теорией эволюции. Действительно, если представить себе систему, непрерывно снабжаемую энергией, в которой некоторые конфигурации способны лучше воспринимать и использовать поступающую энергию, чем другие. Эти последние постепенно исчезают вследствие диссипации, потери ими энергии, в то время, как первые могут компенсировать свои потери и даже расти. Такие рассуждения повторяют ход рассуждений Дарвина и свидетельствуют о том, что дарвиновский принцип выживания приспособленных применим не только к биологической эволюции.

С точки зрения математики самоорганизация состоит в потере или понижении степени симметрии. Например, до кристаллизации жидкость — полностью однородна и изотропна, т.е. полностью симметрична. После замерзания, то есть кристаллизации, в ней появляются выделенные направления, определяемые симметрией кристаллической решетки, тем самым степень симметрии системы понижается.

8.2. Проблема обратимости времени и ее связь с теорией нелинейных систем

8.2.1. Стрела времени и законы динамики. Разговор о тенденциях развития мира имеет смысл начать с того, чтобы разобраться, что такое время, и в первую очередь отказаться от попыток дать точное определение этого понятия, так же, как в геометрии отказались от определения точки или прямой. Время, так же, как и

пространство, является фундаментальным понятием физики, и даже более того, — фундаментальным понятием бытия. Физика нашла способы описания времени, введя его эталон и указав способ приписывания каждому событию некоторого числа, характеризующего момент его наступления.

Если рассматривать время как некоторое измерение нашего мира, то одна из проблем, связанных со временем, состоит в ответе на вопрос об отличии будущего от прошлого. Дело в том, что все законы классической динамики, теории относительности, квантовой физики не содержат в себе отличий прошлого от будущего, в то время, как феноменология, т.е. наблюдаемые процессы в макроскопической физике, химии, биологии, социальные явления, резко разделяют прошлое и будущее.

Действительно, в уравнениях механики, например, можно заставить двигаться все частицы системы по тем же траекториям, но в обратном направлении, или как бы из будущего в прошлое, вполне выполнимой, хотя бы и теоретически, процедурой изменения знаков у импульсов всех частиц, заменой $\mathbf{p}$ на $-\mathbf{p}$, для этого скорости всех частиц должны поменять свое направление. Но этого можно добиться и изменением знаков у времени, т.е. заменой t на $-t$ в уравнениях движения, обращением времени. Таким образом, законы механики не запрещают обратить процессы. То же самое касается и законов электродинамики, и релятивистской и квантовой механики. Однако разбитая ваза не собирается в целую, люди сначала рождаются, живут и умирают, но не наоборот, — эти явления демонстрируют так называемую стрелу времени, указывающую на вполне определенное направление возможного его течения. Как же концепция стрелы времени может возникнуть из фундаментальных законов физики, обладающих симметрией по времени, какими являются законы классической и квантовой механики, электродинамики и др.?

Вплоть до первой половины XIX века обратимость времени считалась идеальным выражением объективности науки, а необратимость явлений жизни объявлялась следствием субъективности наших ощущений. Однако с рассмотрением таких процессов, как диффузия, вязкость и т.п. в науку пришло представление о том, что необратимость физических явлений — это фундаментальное свойство нашего мира, и эта точка зрения теперь подкрепляется развитием современных представлений о нелинейных процессах с хаосом.

Хаотические системы являются неустойчивыми, однако эта неустойчивость может играть и конструктивную роль, которая может помочь в разрешении трудностей современной научной концепции. В рамках не только классических разделов физики, но и в современных ее разделах, где время входит симметрично, основная трудность состоит в понимании и объяснении процессов возникновения нового,

того, чего раньше не существовало, поскольку возникновение нового нарушает симметрию физики: в полностью симметричном мире с обратимым временем и законы эволюции должны быть симметричны.

Для объяснения этих трудностей еще в античности философы вводили понятие случайного воздействия, возмущающего детерминированное поведение детерминированных систем. Примером этого служит "клинамен" Лукреция, который был введен им для описания падения атомов в пустоте. К таким же по сути идеям пришел и Эйнштейн при описании спонтанного испускания света возбужденным атомом. Такие же соображения случайностей привлекаются при описании термодинамических неравновесных процессов.

Эти идеи, по сути, требуют основательного пересмотра фундамента физики, поскольку случайность не свойственна описанию явлений ни для классической механики, ни для электродинамики, ни для теории относительности и т.п. Главную роль в формировании современных взглядов на основы физики могут играть идеи неустойчивости и хаоса.

Действительно, законы классической механики просты и обратимы только для относительно простых систем — типа одной планеты, вращающейся вокруг Солнца, т.е. для так называемых вполне интегрируемых систем. Однако решение даже задачи трех тел уже демонстрирует хаотическое поведение, характеризующееся неустойчивостью; формальная сторона этих проблем была рассмотрена нами в разделе 4.3.4, когда мы говорили о КАМ-теории. В настоящее время стало ясно, что подавляющее большинство систем не являются вполне интегрируемыми, и регулярность поведения систем даже в классической механике — скорее, экзотический частный случай, чем правило. Хаос же обладает неустойчивостью, исключительно чувствительной к сколь угодно малым возмущениям. Абсолютно точные наблюдения нам недоступны, и поскольку это утверждение имеет принципиальный, фундаментальный характер, то следовательно, описание системы в терминах траекторий является чрезмерной идеализацией, и конструктивная роль хаоса проявляется в том, что свойства хаотических систем требуют создания нового описания физических систем, не сводящееся к описанию в терминах траекторий и отражающее свойства динамического хаоса. К такому выводу пришел И. Пригожин, который предложил такое несводимое описание, и на основе его попытался объяснить фундаментальный характер стрелы времени.

Ныне общеприняты две формулировки законов физики — первая основана на понятиях траекторий для классической механики или волновых функций для квантовой. Вторая основана на теориях ансамблей Гиббса и Эйнштейна; вторая формулировка, с точки зрения И. Пригожина, не вносит принципиально нового в динамику, посколь-

ку, будучи применена к отдельной траектории или волновой функции, сводится к первой, обладающей временной симметрией. Описание И. Пригожина — несводимо, так как не может быть применено к отдельной траектории, траектория, по мнению И. Пригожина, не только ненаблюдаема, но и принципиально не пригодна для описания физики явлений, обладающих стрелой времени. Он отказывает траекториям в праве на существовании и формулирует новую динамику, которая приводит к результатам, недостижимым ни в рамках ньютоновой физики, ни в рамках квантовой механики.

8.2.2. Квантовомеханический и космологический парадоксы. Говоря о стреле времени, нельзя не упомянуть еще о двух парадоксах современной физики, тесно с ней связанных, — о квантовомеханическом и космологическом. В квантовой механике фундаментальное описание проводится в терминах волновой функции ψ, и принципиальное различие между классической динамикой и квантовой физикой состоит в том, что классические траектории соответствуют наблюдениям, а волновые функции — соответствуют амплитудам вероятности, и чтобы получить сами вероятности, необходимо дополнительное действие — "коллапс" волновой функции, состоящий в переходе от ψ к $|\psi|^2 = \psi\psi^*$. Это описание не входит в фундаментальное описание квантовомеханической системы, которым является уравнение Шрёдингера.

Природа коллапса до сих пор является предметом многих споров. Многие физики считают, что ответственность за коллапс несет наблюдатель и проводимые им измерения. В этом и заключается квантовомеханический парадокс, который состоит во введении субъективного элемента — наблюдателя — в фундаментальное описание природы.

Трудности, возникающие в связи с такой точкой зрения, состоят в том, что природа не может проявиться без наблюдателя. Об этом говорит знаменитый пример Шрёдингера с кошкой, в котором рассматривается следующий мысленный эксперимент. В закрытую комнату помещают кошку и квантовый объект — например, радиоактивный атом, распад которого включает устройство, убивающее кошку. Согласно сформулированной концепции, до наблюдения кошка находится в смешанном состоянии — "между жизнью и смертью", наблюдение же приводит к коллапсу волновой функции, так как только глядя на кошку, можно сказать, жива она или мертва. В этой ситуации наблюдатель, увидевший мертвую кошку, несет ответственность за ее гибель, так как до наблюдения она не была "полностью мертва".

Коллапс волновой функции связывают также с существованием двух волн амплитуды вероятности — ψ и ψ^*, одна из которых движется из прошлого в будущее, а другая — из будущего в прошлое, и в настоящем происходит перемножение этих волн — коллапс.

Третий, космологический парадокс, связан с асимметрией времени из-за конечного времени существования Вселенной. В традиционном понимании природы между прошлым и будущим нет никакой разницы, и в этом проявляется объективность законов природы; наличие стрелы времени объясняется при этом нашим субъективным восприятием. Однако современные космологические представления приписывают нашей Вселенной возраст порядка 15 млрд. лет, прошедших после ее рождения в результате Большого Взрыва. По терминологии И. Пригожина, этот Взрыв был “событием”, т.е. таким явлением, наступление которого вовсе не обязательно, но которое повлекло за собой существенное изменение истории. В классической формулировке законов физики понятие события отсутствует.

Свойства хаотических динамических систем, изученные нами в главе 5, лежат в основе пути, на котором, возможно, будет достигнуто решение этих парадоксов. Этот путь предложен И. Пригожиным, и для того, чтобы понять, в чем его отличие от подходов Гиббса и Больцмана, описывающих системы в терминах статистических ансамблей, рассмотрим более детально, как необратимость времени возникает в уравнениях физики при переходе от описаний в терминах траекторий к описанию в терминах ансамблей.

8.2.3. Причина необратимости времени в статистической физике. Исторически первый пример, в котором возникает уравнение, лишенное симметрии во времени, — это уравнение Больцмана, известное в статистической физике, следствием которого является закон возрастания энтропии в неравновесных изолированных системах. Сам Больцман считал необратимость времени фундаментальным свойством природы, несмотря на резкую критику этой концепции со стороны его современников, и в частности, Лошмидта. Больцман называл XIX век веком эволюционной теории Дарвина, полагая, что эволюция, основанная на принципе естественного отбора, является всеобщим принципом развития (такой отбор, по существу, является событием в понимании И. Пригожина), и, по-видимому, надеялся, что описание в терминах ансамблей поможет вывести законы эволюции сложных систем “на кончике пера”.

В статистической теории открытых систем энтропия является одной из важнейших характеристик и служит мерой неопределенности при статистическом описании, мерой упорядоченности неравновесных состояний и мерой разнообразия в теории эволюции. В зависимости от детальности описания можно использовать то или иное определение энтропии. Если система состоит из N частиц, то для динамического микроскопического описания системы в 6-мерном

пространстве с координатами $x = (\mathbf{r}, \mathbf{p})$ достаточно задать функцию

$$N(x,t) \equiv \sum_{i=1}^{N} \delta(x - x_i(t)), \quad \int N(x,t)dx = N.$$

Интеграл

$$\int\limits_{\Delta\mathbf{p}} \int\limits_{\Delta\mathbf{r}} N(x,t)d\mathbf{r}d\mathbf{p}$$

равен числу частиц, находящихся в момент времени t в окрестности $\Delta\mathbf{r} \subset \mathcal{R}_3$ точки $\mathbf{r} \in \mathcal{R}_3$ и имеющих импульс из окрестности $\Delta\mathbf{p} \subset \mathcal{R}_3$ точки $\mathbf{p} \in \mathcal{R}_3$. Функция $N(x,t)$ не несет информации о взаимосвязи между координатами и скоростями разных частиц, и при статистическом описании рассматривается как случайная функция. Для задания пространства ее значений (аналог пространства элементарных событий в теории вероятности) вводят ансамбль Гиббса как множество систем с одними и теми же макроскопическими параметрами — объемом, средней плотностью, энергией (температурой), — но отличающихся микроскопическими параметрами. Все микроскопические состояния считаются равноправными (равновероятными). Усреднение по ансамблю Гиббса будем обозначать угловыми скобками: $\langle\cdot\rangle$:

$$\langle N(x,t)\rangle = \left\langle \sum_{i=1}^{N} \delta(x - x_i(t)) \right\rangle.$$

Если $f(x,t)$ — плотность вероятности координат частицы, то в силу линейности математического ожидания,

$$\langle N(x,t)\rangle = Nf(x,t).$$

Энтропия Больцмана равна

$$S_B(t) = -kn \int f(x,t) \ln f(x,t)dx + S_0,$$

где S_0 — несущественная константа, $n = N/V$ — средняя плотность частиц.

Если рассмотреть $6N$-мерное пространство координат и импульсов частиц, обозначив $X = (\mathbf{r}_1, \ldots, \mathbf{r}_N, \mathbf{p}_1, \ldots, \mathbf{p}_N)$, то микроскопическое динамическое поведение системы N частиц определится функцией

$$f_N^d(X,t) = \delta(X - X(t)), \quad \int f_N^d(X,t)dx = 1. \tag{8.1}$$

В отличие от функции $N(x,t)$, $x \in \mathcal{R}_1$, описывающей поведение "средней" частицы, функция $f_N^d(x,t)$ несет информацию о движении всех N частиц одновременно. Усреднение по ансамблю приведет нас к динамическому распределению

$$f_N(X,t) = \langle \delta(X - X(t)) \rangle,$$

и энтропии Гиббса

$$S_G(t) = -kn \int f_N(X,t) \ln f_N(X,t) dX + S_0'. \tag{8.2}$$

В отличие от предыдущего описания, в энтропии Гиббса учтены все корреляции между положениями и импульсами частиц, поэтому

$$S_G \leq S_B.$$

Система, состоящая из большого числа частиц, обладает свойствами системы с хаотическим поведением, т.е. неустойчивостью по отношению к малым возмущениям состояния системы, поэтому уравнения для микроскопических переменных $N(x,t)$, $f_N^d(X,t)$, эквивалентные уравнениям Гамильтона и симметричные относительно обращения времени, не могут дать удобного аппарата описания состояний системы, а служат лишь основой для перехода к макроскопическому описанию. Этот запрет к использованию обратимых уравнений для описания динамики системы становится фундаментальным свойством больших систем, если принять за основу принципиальную невозможность абсолютно точного задания координат и импульсов каждой частицы системы.

Неустойчивость и чувствительность к возмущениям микроскопического динамического описания больших систем не позволяет также сразу приступить к усреднению уравнений (8.1) для того, чтобы действовать более аккуратно, сначала перейдем к "сглаженному" описанию системы, вводя ансамбль Гиббса таких систем, в которых предварительно произведено усреднение состояний частиц по физически бесконечно малому объему. Для этого, следуя Ю.Л.Климонтовичу, введем понятие физически бесконечно малой длины l_{ph} и времени τ_{ph} как величин, обладающих следующими свойствами:

— величины l_{ph} и τ_{ph} малы по сравнению с характерными масштабами (пространственными и временными) рассматриваемой системы;

— физически бесконечно малый объем $(l_{ph})^3$ содержит достаточно большое число частиц $N_{ph} = nV_{ph} \gg 1$.

Проведем усреднение микроскопических состояний системы по физически бесконечно малому объему и введем тем самым сглаженные распределения

$$\widetilde{f}_N(X,t) = \langle f_N^d(X,t) \rangle_{V_{ph}},$$

$$\widetilde{N}(x,t) \equiv n\widetilde{f}(x,t) = \langle N(x,t)\rangle V_{ph}.$$

Для макроскопических систем сглаженные распределения $\widetilde{f}_N(X,t)$, $\widetilde{N}(x,t)$ — случайные функции, поэтому возможно дальнейшее усреднение по ансамблю Гиббса, в результате которых получаются макроскопические функции распределения

$$f_N(X,t) = \langle \widetilde{f}_N(X,t)\rangle,$$

$$nf(x,t) \equiv \langle \widetilde{N}(x,t)\rangle \equiv n\langle \widetilde{f}(x,t)\rangle.$$

Рассмотрим теперь, как производится усреднение по ансамблю Гиббса, и выясним, на каком этапе возникает необратимость.

Пусть динамические уравнения (уравнения Гамильтона) являются абсолютно точными, и вся неопределенность статистического описания, т.е. то, чем отличаются системы ансамбля Гиббса, сосредоточена в начальном моменте. Тогда статистическое распределение связано с микроскопическим динамическим следующим равенством:

$$f_N(X,t) = \int \delta(X - X(X_0, t - t_0)) f_N(X_0) dX_0, \tag{8.3}$$

и уравнение для f_N в этом случае по форме совпадает с динамическими гамильтоновыми уравнениями:

$$\frac{df_N}{dt} + [H(X),\, f_N(X,t)] = 0, \tag{8.4}$$

или, более подробно,

$$\frac{\partial f_N}{\partial t} + \mathbf{v}\frac{\partial f_N}{\partial \mathbf{r}} + \mathbf{F}^m(\mathbf{r},t)\frac{\partial f_N}{\partial \mathbf{p}} + [H(X),\, f_N(X,t)] = 0;$$

здесь $[\cdot,\,\cdot]$ — скобки Пуассона:

$$[a,\, b] = \sum_{i=1}^{3N} \frac{\partial a}{\partial p_i}\frac{\partial b}{\partial r_i} - \frac{\partial a}{\partial r_i}\frac{\partial b}{\partial p_i},$$

$H(X)$ — гамильтониан, а микроскопическая сила $\mathbf{F}^m(\mathbf{r},t)$ дается формулой

$$\mathbf{F}^m(\mathbf{r},t) = -\frac{\partial}{\partial \mathbf{r}} \int \Phi(|\mathbf{r} - \mathbf{r}'|) N(\mathbf{r}', \mathbf{p}', t) d\mathbf{r}' d\mathbf{p}',$$

в которой $\Phi(|\mathbf{r} - \mathbf{r}'|)$ — потенциал взаимодействия частиц.

Уравнение (8.4) называется уравнением Лиувилля и обратимо, так как в нем неопределенность просто переносится по траекториям. Обратимость эта проявляется, в частности, в том, что энтропия Гиббса (8.2) при эволюции, описываемой уравнением (8.4), остается неизменной:

$$\frac{dS_G}{dt} = 0.$$

Откуда же может появиться необратимость? Существенную роль в уравнениях (8.3), (8.4) играют взаимодействие и столкновения частиц, при эволюции системы именно они являются источниками неустойчивости и хаоса, чувствительного к возмущениям. При усреднении же по физически бесконечно малому объему, проводимому перед операцией усреднения по ансамблю Гиббса, теряется информация о корреляциях между скоростями и положениями частиц, возникающих на физически бесконечно малых масштабах в результате этих взаимодействий. Усреднение по физически бесконечно малому объему приводит, с одной стороны, к тому, что неконтролируемые малые "внешние воздействия" и принципиально неустранимые "размытости" в задании траекторий частиц, имеющие микроскопические масштабы, т.е. меньшие, чем физически бесконечно малые, усреднятся, и описание системы после такого усреднения можно считать не зависящим от микрофлуктуаций. С другой стороны, усреднение приводит к сглаживанию, и хотя хаотичное поведение системы может сохранится, но масштабы возмущений, к которым будет чувствительна "сглаженная" система, превосходит физически бесконечно малые. Такое усреднение может трактоваться как следствие невозможности проводить абсолютно точные измерения координат частиц и времени событий.

Уравнение Больцмана получено для "одночастичной" функции $f(x,t) \equiv f(\mathbf{r},\mathbf{p},t)$, описывающей среднюю по ансамблю Гиббса плотность частиц в заданной точке $x = (\mathbf{r},\mathbf{p})$ в момент времени t. Замкнутое уравнение для $f(x,t)$ можно получить из (8.3), (8.4), пренебрегая корреляциями между состояниями частиц $(\mathbf{r},\mathbf{p})$ и $(\mathbf{r}',\mathbf{p}')$. При этом можно пренебрегать корреляциями всех масштабов, записав для усредненных по ансамблю Гиббса плотностей

$$f_2(x,x',t) = f(x,t)f(x',t), \tag{8.5}$$

что является достаточно грубым упрощением, так как изменение состояния происходит за счет взаимодействий частиц, и прошлое от будущего отличается тем, что частицы "помнят" о своих взаимодействиях, что и отражается в корреляциях. Более мягким упрощением является пренебрежение только мелкомасштабными корреляциями, что выражается в равенстве для случайных функций, усредненных

по физически бесконечно малому объему:

$$\widetilde{f}_2(x, x', t) = \widetilde{f}(x, t)\widetilde{f}(x', t). \tag{8.6}$$

Действительно, при усреднении (8.6) по ансамблю Гиббса крупномасштабные корреляции сохранятся, а мелкомасштабные пропали при усреднению по физически бесконечно малым объемам и временам.

Заметим, что соотношение (8.5) является точным для динамических микроскопических описаний:

$$f_2^{\partial}(x, x', t) = f^{\partial}(x, t) f^{\partial}(x', t),$$

и приводят к обратимым уравнениям. При использовании же условий (8.6) тем или иным способом приходят к уравнению Больцмана, имеющего вид

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}} + \mathbf{F}(\mathbf{r}, t)\frac{\partial}{\partial \mathbf{p}}\right) f(\mathbf{r}, \mathbf{p}, t) = I_B(\mathbf{r}, \mathbf{p}, t), \tag{8.7}$$

$$V^{-1}\int f d\mathbf{r} d\mathbf{p} = 1,$$

где

$$\mathbf{F} = \mathbf{F}_0 - n\frac{\partial}{\partial \mathbf{r}}\int \Phi(|\mathbf{r} - \mathbf{r}'|) f(\mathbf{r}', \mathbf{p}', t) d\mathbf{r}' d\mathbf{p}', \tag{8.8}$$

$\mathbf{F}_0$ — внешняя сила, а второе слагаемое в (8.8) — средняя сила, определяемая одноточечным распределением $f(\mathbf{r}, \mathbf{p}, t)$; такого рода силы рассматривались А.А.Власовым и носят его имя.

Правая часть (8.7) называется интегралом столкновений, и учитывает увеличение и уменьшение плотности распределения за счет той части взаимодействий, которая приводит к диссипации. В разреженном газе такие процессы в основном обусловлены столкновениями частиц, этим и объясняется название $I_B(\mathbf{r}, \mathbf{p}, t)$. Если $I_B(\mathbf{r}, \mathbf{p}, t) = 0$, то уравнение (8.7) будет обратимым во времени. Л.Больцман получил интеграл столкновений в виде

$$\begin{gathered} I_B(\mathbf{r}, \mathbf{p}, t) = n\int_0^{\infty} \rho d\rho \int_0^{2\pi} d\varphi \int d\mathbf{p}_2 |\mathbf{v} - \\ - \mathbf{v}_2| [f(\mathbf{r}, \mathbf{p}', t) f(\mathbf{r}, \mathbf{p}_2', t) - f(\mathbf{r}, \mathbf{p}, t) f(\mathbf{r}, \mathbf{p}_2, t)], \end{gathered} \tag{8.9}$$

где $\mathbf{p}'$, $\mathbf{p}_2'$ — импульсы частиц до (или после) столкновений, выраженные через $\mathbf{p}$ и $\mathbf{p}_2$, используя законы сохранения импульса и энергии:

$$\mathbf{p} + \mathbf{p}_2 = \mathbf{p}' + \mathbf{p}_2', \quad \frac{\mathbf{p}^2}{2m} + \frac{\mathbf{p}_2^2}{2m} = \frac{(\mathbf{p}')^2}{2m} + \frac{(\mathbf{p}_2')^2}{2m},$$

и при интегрировании в (8.9) ось z направлена вдоль вектора относительной скорости частиц $\mathbf{v} - \mathbf{v}_2$.

Интеграл столкновений обладает следующими свойствами:

$$\int \psi(\mathbf{p}) I_B(\mathbf{r}, \mathbf{p}, t) d\mathbf{p} = 0,$$

если $\psi(\mathbf{p}) = 1$, или $\psi(\mathbf{p}) = \mathbf{p}$, или $\psi(\mathbf{p}) = \frac{\mathbf{p}^2}{2m}$, и

$$\int \psi(\mathbf{p}) I_B(\mathbf{r}, \mathbf{p}, t) d\mathbf{p} \geq 0,$$

если $\psi(\mathbf{p}) = -k \ln f(\mathbf{r}, \mathbf{p}, t)$, откуда следует, что для решения уравнения (8.7) выполнены законы сохранения числа частиц, импульса и энергии, но энтропия — возрастает.

Итак, формально необратимость при описании систем, состоящих из большого числа частиц, возникает при усреднении по физически бесконечно малым объемам и временам, что приводит к потере информации о корреляциях между положениями и импульсами частиц на масштабах, неразличимых для экспериментатора.

8.2.4. Описание движения, несводимое к траекториям. Обратимся теперь к операторному описанию ансамблей в классической и квантовой механике. Введем гильбертово пространство $\mathcal{R}$, элементом которого будем считать плотность вероятности f, и запишем уравнение Лиувилля (8.4) формально в операторном виде

$$i\frac{df}{dt} = \mathcal{L}f, \tag{8.10}$$

где $\mathcal{L}$ — линейный оператор Лиувилля, действующий в гильбертовом пространстве $\mathcal{R}$, и рассмотрим задачу на собственные значения оператора Лиувилля:

$$\mathcal{L}\varphi_n = l_n \varphi_n, \quad \varphi_n \in \mathcal{R}.$$

Так как оператор $\mathcal{L}$ — эрмитов, то его собственные значения l_n вещественны, а $\{\varphi_n\}$ можно выбрать так, чтобы они образовывали полный ортонормированный базис $\mathcal{R}$. Разложим f по базису $\{\varphi_n\}$:

$$f = \sum_{n=1}^{\infty} c_n \varphi_n,$$

и подставим в (8.10). Эволюция во времени f теперь определится соотношением вида

$$f(t) = U(t) f(0) = e^{-i\mathcal{L}t} f(0). \tag{8.11}$$

Как и в квантовой механике, $e^{-i\mathcal{L}t}$ — унитарный оператор, обладающий свойством изометричности, т.е. сохранения нормы, поэтому эволюцию элемента $f \in \mathcal{R}$,

$$f(t) = \sum_{n=1}^{\infty} c_n e^{il_n t} \varphi_n,$$

можно представить как вращение мод φ_n в гильбертовом пространстве $\mathcal{R}$. Оператор вращения обратим, а для того, чтобы ввести необратимость, выделенность направления времени, необходимо ввести затухание, что можно сделать формально, рассматривая комплексные значения собственных чисел l_n оператора Лиувилля, однако это невозможно из-за эрмитовости $\mathcal{L}$.

Таким образом, для того, чтобы достичь необратимости, надо либо отказаться от уравнения Лиувилля и выбрать некоторое другое, необратимое уравнение, как это сделал Л.Больцман, либо отказаться от метрики гильбертова пространства и ввести новую метрику. По второму пути пошел И. Пригожин, предложив новое описание динамических систем в терминах оснащенных гильбертовых пространств.

Заметим, что получить выражения в явном виде для собственных значений и собственных функций оператора $\mathcal{L}$ можно далеко не всегда, а лишь для случая интегрируемых систем. Если же применять методы теории возмущения, представляя оператор Лиувилля в виде $\mathcal{L} = \mathcal{L}_0 + \lambda\mathcal{L}_1$, где $\mathcal{L}_1$ — некоторое возмущение оператора Лиувилля $\mathcal{L}_0$ с известным спектром, то следует представить собственные функции оператора $\mathcal{L}$ в виде суммы по степеням λ слагаемых вида $1/(l_n - l_m)$, и при

$$|l_n - l_m| \ll 1 \tag{8.12}$$

возникает известная проблема малых знаменателей (с этой проблемой мы уже сталкивались, когда рассматривали асимптотические методы Крылова–Боголюбова решения возмущенных нелинейных задач теории колебаний). Эта проблема особенно остро встает для так называемых больших систем Пуанкаре, характеризуемых тем, что в них соотношение (8.12) возникает неизбежно. Это свидетельствует в пользу того, чтобы попробовать найти иное описание систем, не сводимое к уравнениям (8.10).

Второе замечание состоит в том, что те же самые слова, которые говорились для уравнения Лиувилля, можно произнести и для уравнения Шрёдингера, описывающего квантовомеханические системы, так как уравнение Шрёдингера совпадает с (8.10) при замене оператора $\mathcal{L}$ на оператор Гамильтона $\mathcal{H}$. При описании ансамблей квантовомеханических систем также используются схожие рассуждения, только оператор Лиувилля будет действовать не на элемент f гильбертова

пространства, а на оператор плотности, и потому называется супероператором.

Вернемся к несводимому описанию динамических систем и рассмотрим подход И. Пригожина на достаточно простых примерах.

Как уже говорилось, для хаотических динамических систем понятие траектории являются слишком сильной идеализацией, и потому рассматривается описание систем в терминах распределений вероятности найти систему в том или ином состоянии. В частности, при описании сдвига Бернулли $\rho(x)$ — вероятность найти систему в состоянии x, и в силу определения преобразования Бернулли для $\rho(\cdot)$ можно записать соотношение

$$\rho_{n+1} = U\rho_n, \tag{8.13}$$

называемое уравнением Перрона–Фробениуса (см. раздел 5.2.2).

Если в терминах квантовой механики или статистической механики эволюции описываются оператором типа $e^{i\mathcal{H}t}$ или $e^{i\mathcal{L}t}$ для квантового и классического описания соответственно, то в теории хаотических ансамблей оператор эволюции должен описывать приближение системы к равновесию, и следовательно, не может быть унитарным, а должен содержать релаксацию. Действительно, в случае сдвига Бернулли оператор U определяется соотношением

$$\rho_{n+1}(x) = U\rho_n(x) \equiv \frac{1}{2}\left(\rho_n\left(\frac{x}{2}\right) + \rho_n\left(\frac{x+1}{2}\right)\right).$$

Для функции $\rho_0(x) = \mathrm{const}$ выполняется соотношение $U\rho_0 = \rho_0$, т.е. распределение вероятности, описываемое элементом $\rho_0(x) = \mathrm{const}$, является равновесным, и достигается при $t \to \infty$. Иными словами, остальные моды, если они существуют, должны затухать. Действительно, оператор U имеет набор собственных функций и отвечающий им набор собственных значений, меньших единицы. Собственные функции — это так называемые полиномы Бернулли, первые несколько полиномов имеют вид

$$B_1(x) = x - \frac{1}{2}, \quad UB_1 = \frac{1}{2}B_1;$$

$$B_2(x) = x^2 - x + \frac{1}{6}, \quad UB_2 = \frac{1}{2^2}B_2;$$

....

Для любого натурального n выполнено равенство

$$UB_n = \frac{1}{2^n}B_n.$$

Операторная формулировка динамики систем, адекватно отражающую их эволюцию, состоит в релаксации всех мод, за исключением тех, которые соответствуют единичному собственному значению оператора U. Задание функции распределения $f(x,t)$ при движении одной частицы формально должно приводить к обобщенной функции $f(x,t) = \delta(x - x_0(t))$, где $x_0(t)$ задает траекторию частицы. Покажем, что описание эволюции системы в виде (8.13) не годится для траекторий, задаваемых в виде обобщенных функций.

Заметим, что U — не унитарный и не самосопряженный оператор, и сопряженный к нему U^+ не имеет собственными функциями полиномы Бернулли, кроме $B_0 = \text{const}$; более того, он не имеет иных собственных функций (кроме B_0) в классе непрерывных функций. Пользуясь его определением, можно записать, что

$$U^+g(x) = \begin{cases} g(2x), & \text{если} \quad 0 \le x < 1/2, \\ g(2x-1), & \text{если} \quad 1/2 \le x \le 1. \end{cases}$$

Можно заметить, что, однако, в классе обобщенных функций у него есть собственные элементы, например,

$$U^+(\delta(x-1) + \delta(x)) = \frac{1}{2}(\delta(x-1) + \delta(x)),$$

и существует целое семейство собственных функций $\widetilde{B}_n(x)$ оператора U^+, соответствующих собственным значениям $1/2^n$.

Таким образом, мы имеем биортогональный базис в гильбертовом пространстве, в котором действует оператор эволюции U — один из них составлен из полиномов Бернулли, а второй — из собственных векторов оператора U^+. Как и в квантовой механике, разложим плотность вероятности $\rho(\cdot)$ по биортогональным базисам $\{B_n\}$ и $\{\widetilde{B}_n\}$,

$$\rho = \sum_{n=1}^{\infty} B_n(\widetilde{B}_n, \rho), \tag{8.14}$$

и рассмотрим, нельзя ли расширить область определения оператора эволюции U, так, чтобы включить в описание динамики как можно более широкий класс функций — эта процедура носит название операции оснащения гильбертова пространства. Однако так как $\widetilde{B}_n$ — обобщенные функции, то ρ в (8.14) не может быть обобщенной. Это и отражает тот фундаментальный факт, что траектории, описывающиеся обобщенными функциями, не могут использоваться для описания динамики системы: разложение (8.14) для них невозможно.

На основе приведенных здесь соотношений И. Пригожин предлагает новую формулировку динамики. Эта формулировка приводит к

выделению направления во времени и отражает тот фундаментальный факт, что описание в терминах траекторий не является адекватным для систем с неустойчивостью, чувствительной к малым возмущениям, характерной для хаотической динамики, [11].

В заключение отметим еще один важный аспект понятия времени, который связан с невозможностью измерения в точно фиксированный момент времени — факт, который при выводе уравнения Больцмана привел нас к необходимости усреднения по физически бесконечно малому интервалу времени. Если принять этот факт за фундаментальный принцип, то любое наблюдение мира является наблюдением картины, усредненной за некоторый промежуток времени, причем для разных наблюдателей этот промежуток оказывается различным. Отсюда следует невозможность полного, абсолютного воспроизведения результатов эксперимента, и в связи с этим возникает фундаментальная проблема формулировки физических законов, инвариантных по отношению к наблюдениям в “размытые” моменты времени.

Этот раздел не столько говорит о том, каких успехов достигла наука в описании динамических систем, сколько демонстрирует основные трудности современной физики и лишь намечает подходы к их преодолению. Поэтому в конце него не следует ставить точку, здесь уместнее многоточие...

8.3. Самоорганизация в активных средах

Вернемся к тенденции усложнения структуры систем и постараемся найти механизмы, лежащие в ее основе.

Самоорганизация как возникновение и усложнение структур в первоначально симметричных средах свойственна нелинейным открытым системам, устойчивые структуры могут возникать, в частности, как результат непрерывного перераспределения энергии, поступающей в систему извне, за счет ее диссипации, переноса путем диффузии или конвекции, а также в результате нелинейного взаимодействия различных компонент среды. Среды, в которых происходят подобные процессы, называют диссипативными.

Простейшую математическую модель возникновения волн и структур в таких средах можно получить, рассматривая среду как совокупность нелинейных элементов — бистабильных (т.е. имеющих два устойчивых состояния), возбудимых, которые после вынужденного выхода из состояния покоя после некоторого цикла превращений вновь возвращаются в исходное устойчивое к малым возмущениям состояние, или автоколебательных. В этом разделе изучаются основные особенности моделей распределенных систем, составленных из

соединенных между собой взаимодействующих элементарных ячеек, обладающих названными характеристиками; они также называются моделями активных сред. Используемое описание близко к моделям клеточных автоматов.

8.3.1. Бистабильные среды. Так называются среды, составленные из совокупности соединенных между собой активных бистабильных элементов. Бистабильные, или триггерные, элементы обладают двумя устойчивыми состояниями, в каждом из которых элемент может находиться достаточно долго. Внешнее воздействие может привести к переходу от одного состояния к другому; для этого необходимо, чтобы воздействие было достаточно велико и превосходило некоторый порог.

1	2	2	2	2	2	2	2	2	2	2	2
1	1	2	2	2	2	2	2	2	2	2	2
1	1	1	2	2	2	2	2	2	2	2	2
1	1	1	1	2	2	2	2	2	2	2	2
1	1	1	1	1	2	2	2	2	2	2	2

Рис. 8.1. Эволюция цепочки из бистабильных элементов

Составим цепочку из активных бистабильных элементов и будем считать, что на их состояние может влиять лишь соседний элемент, причем это влияние происходит тогда и только тогда, когда соседние элементы находятся в разных состояниях. Пусть исходное состояние всех элементов — одно и то же, и при возбуждении крайнего элемента он переходит из исходного метастабильного состояния в другое, абсолютно стабильное, и принуждает к этому своего соседа. В результате по цепочке распространяется волна переключения (рис. 8.1).

Это описание проведено в терминах так называемых клеточных автоматов и очень удобно для компьютерного моделирования. Более детальное описание основано на дифференциальных уравнениях в частных производных; в этом случае говорят о средах с непрерывно изменяющимися (распределенными) параметрами.

Математическая модель бистабильного элемента. Рассмотрим математическую модель бистабильного элемента. Этот элемент может быть описан как некоторая система, обладающая двумя положениями устойчивого равновесия, например, описываемая уравнением типа

$$\dot{u} = f(u), \qquad (8.15)$$

где u — некоторая переменная, описывающая состояние системы, а функция $f(\cdot)$ имеет вид, изображенный на рис. 8.2.

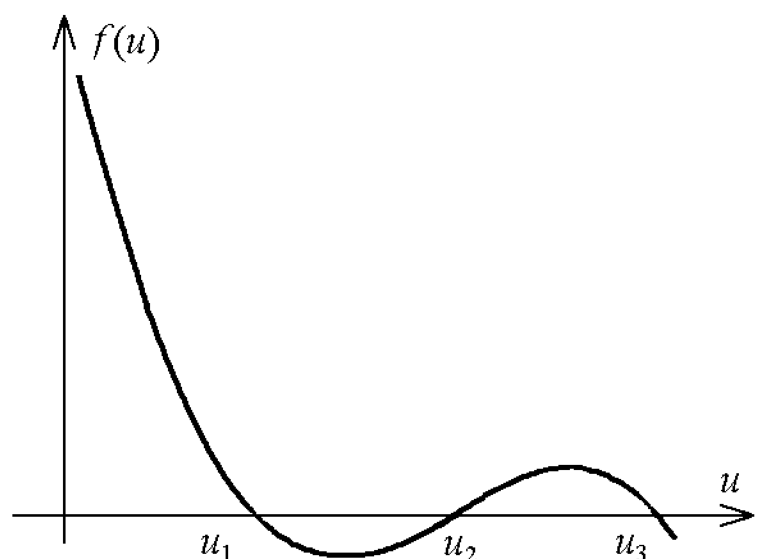


Рис. 8.2. График правой части уравнения, описывающего бистабильный элемент

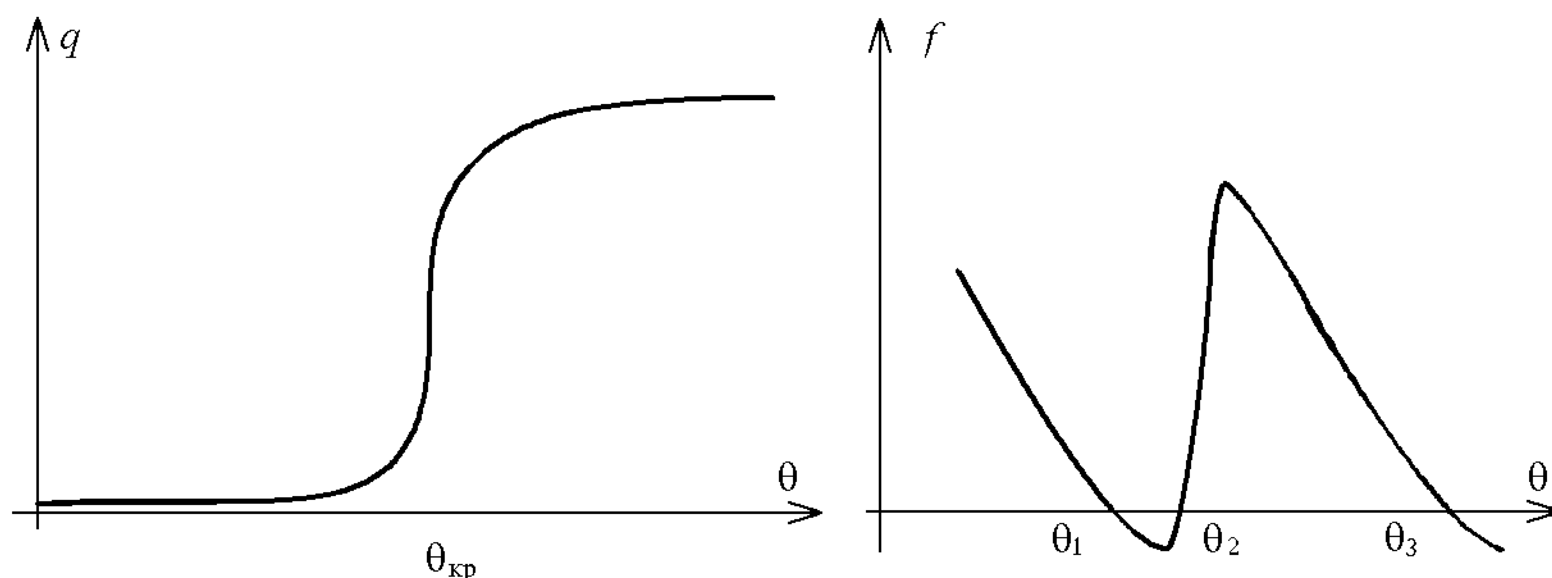


Рис. 8.3. Количество выделяющегося тепла в зависимости от температуры: (*а*) – без теплообмена со средой, (*б*) – при наличии теплообмена

Действительно, стационарное состояние элемента дается равенством $f(u) = 0$, а устойчивость его определяется знаком производной $f'(u)$ так, что при расположении корней функции $f(\cdot)$, изображенном на рис. 8.2, элемент, описываемый соотношением (8.15), обладает двумя устойчивыми состояниями $u = u_1$ и $u = u_3$, а стационарное состояние $u = u_2$ является неустойчивым.

Бистабильный элемент можно построить на основе физических законов, например, как ячейку, в которой температура ϑ может изменяться за счет выделения тепла при горении и за счет теплообмена с окружающей средой. Для создания ячейки с двумя устойчивыми состояниями равновесия положим, что при температуре ϑ, меньшей некоторой критической, горение в ячейке не возникает, а при $\vartheta > \vartheta_{\text{кр}}$ протекает реакция горения с выделением тепла, количество q_f которого при $\vartheta \gg \vartheta_{\text{кр}}$ не зависит от ϑ (график зависимости $q_f(\vartheta)$ приведен на рис. 8.3,*а*). Теплообмен со средой будем описывать законом Фурье, согласно которому количество поступающего в ячейку тепла пропорционально разности температур ячейки ϑ и среды ϑ_{env}. Из уравнения теплового баланса получим, что температура ячейки описывается уравнением

$$\dot{\vartheta} = \frac{1}{C}(q_f(\vartheta) - \gamma(\vartheta - \vartheta_{\text{env}})), \tag{8.16}$$

где C — теплоемкость ячейки, γ — коэффициент теплообмена. График зависимости правой части уравнения (8.16) от ϑ приведен на рис. 8.3,*б*.

Заметим, что в зависимости от параметров γ, ϑ_{env}, $\vartheta_{\text{кр}}$ возможны различные расположения стационарных точек, и даже различное их число. Так, в ситуации, изображенной на рис. 8.4,*а*, существует лишь одно устойчивое положение равновесия, при котором отсутствует горение и теплообмен с окружающей средой, поскольку выполнены равенства $\vartheta = \vartheta_1 = \vartheta_{\text{env}}$, означающие, что равновесие достигнуто при

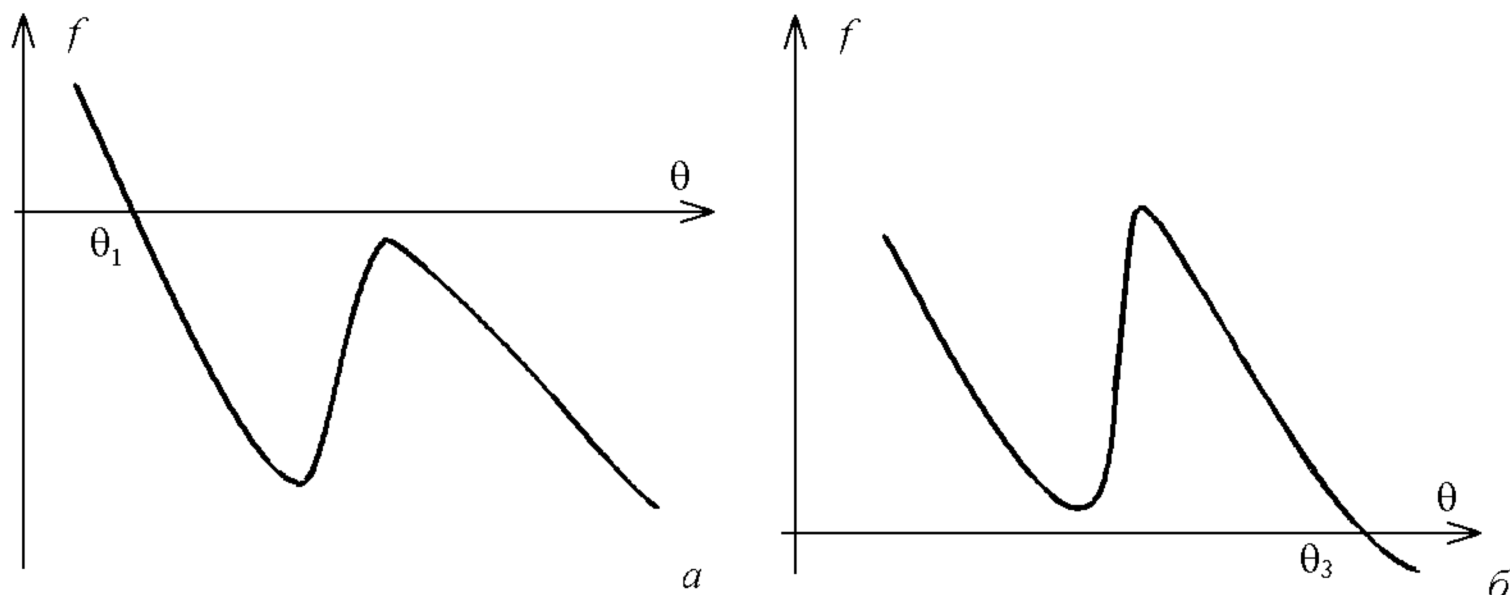
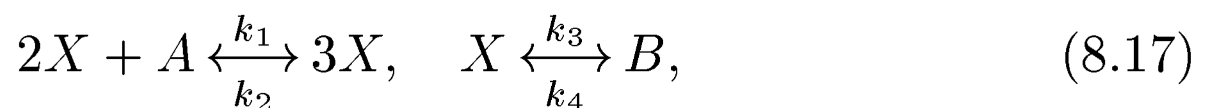


Рис. 8.4. Количество тепла в зависимости от температуры при различных режимах теплообмена: самоподдерживающееся горение невозможно (*а*); состояние устойчивого горения (*б*)

температуре, равной температуре окружающей среды. Это режим, при котором невозможно самоподдерживающееся горение. В ситуации, изображенной в правой части рис. 8.4,*б*, тоже есть единственная устойчивая точка, при которой энергия, выделяющееся при горении, целиком компенсируется теплоотдачей в окружающую среду. Это режим самовозгорания.

Функцию, график которой изображен на рис. 8.2, можно задать полиномом третьего порядка. Такие функции возникают в уравнениях химической кинетики, описывающих временную зависимость концентрации вещества $X(t)$ в реакции, в которой две молекулы вещества X, соединяясь с молекулой вещества A, образуют три молекулы вещества X, само же вещество X превращается в вещество B, причем реакции соединения и распада могут идти как в “прямую”, так и в “обратную” сторону, а концентрации веществ A и B поддерживаются постоянными. Схематически такие реакции записываются в виде соотношений



$$2X + A \underset{k_2}{\overset{k_1}{\longleftrightarrow}} 3X, \quad X \underset{k_4}{\overset{k_3}{\longleftrightarrow}} B, \tag{8.17}$$

где k_i, $i = 1, \ldots, 4$ обозначают скорости реакций. Кинетическое дифференциальное уравнение для концентрации $X(t)$ имеет вид

$$\dot{X} = k_1 A X^2 - k_2 X^3 - k_3 X + k_4 B,$$

и специальным выбором коэффициентов k_i, $i = 1, 2, 3, 4$, и концентраций A и B можно добиться существования двух устойчивых состояний реакции.

В экологических моделях бистабильные системы возникают при моделировании роста популяций. Рождение новой особи происходит в результате встречи мужской и женской особей. Так как число и тех, и других пропорционально объему популяции $n(t)$, то число встреч

пропорционально n^2; количество смертей тоже пропорционально $n(t)$, поэтому без учета диффузионного блуждания особей получаем модель популяции в виде

$$\frac{\partial n}{\partial t} = -\gamma n + \alpha m(n) n^2, \tag{8.18}$$

где γ — коэффициент смертности, $\alpha m(n)$ — коэффициент скорости размножения, пропорциональный массе пищи $m(n)$, который, в свою очередь, зависит от числа поедающих ее особей. В простейшем случае $m(n) = e^{-n/n_0}$, и для $n \ll n_0$ в правой части (8.18) возникает полином третьей степени.

Модель бистабильной среды. Если вернуться к модели горения и соединить ячейки между собой в цепочку, то изменение температуры может происходить и за счет теплообмена между соседними ячейками, и тогда в предположении непрерывной среды получим уравнение

$$\frac{\partial \vartheta}{\partial t} = \frac{1}{C}(q_f(\vartheta) - \gamma(\vartheta - \vartheta_{\text{env}})) + k\frac{\partial^2 \vartheta}{\partial x^2}, \tag{8.19}$$

где k — температуропроводность.

В двумерном и трехмерном случаях к уравнению вида

$$\dot{u} = f(u) + D\Delta u, \tag{8.20}$$

где D — коэффициент диффузии, приводят примеры и из химической кинетики, и из популяционной биологии. Проанализируем его решение. Как мы уже отмечали, основной тип поведения, характерный для бистабильных сред — это волны переключения. Волна переключения из состояния $u = u_3$ в состояние $u = u_1$ описывается соотношением

$$u = u(\xi), \quad \xi = x - ct, \tag{8.21}$$

с граничным условием

$$u \to u_3 \quad \text{при} \quad \xi \to -\infty, \quad u \to u_1 \quad \text{при} \quad \xi \to \infty.$$

Подставив (8.21) в (8.20), получим

$$-cu' = f(u) + Du'', \tag{8.22}$$

где штрих обозначает производную по ξ.

Для того чтобы найти скорость c волны переключения, воспользуемся следующей аналогией. Введем функцию $U(u) = \int\limits_0^u f(z)dz$, тогда $f = \partial U/\partial u$, и вместо (8.22) можем записать

$$Du'' = -\frac{\partial U}{\partial u} - cu'. \tag{8.23}$$

Это уравнение можно интерпретировать как уравнение движения частицы массы D с координатой u в среде с сопротивлением, если рассматривать формально ξ как время, а U — как потенциал некоторого поля. Согласно определению U, при $u = u_1$ и $u = u_3$ функция U достигает локальных максимумов, а при $u = u_2$ — локального минимума, как показано на рис. 8.5.

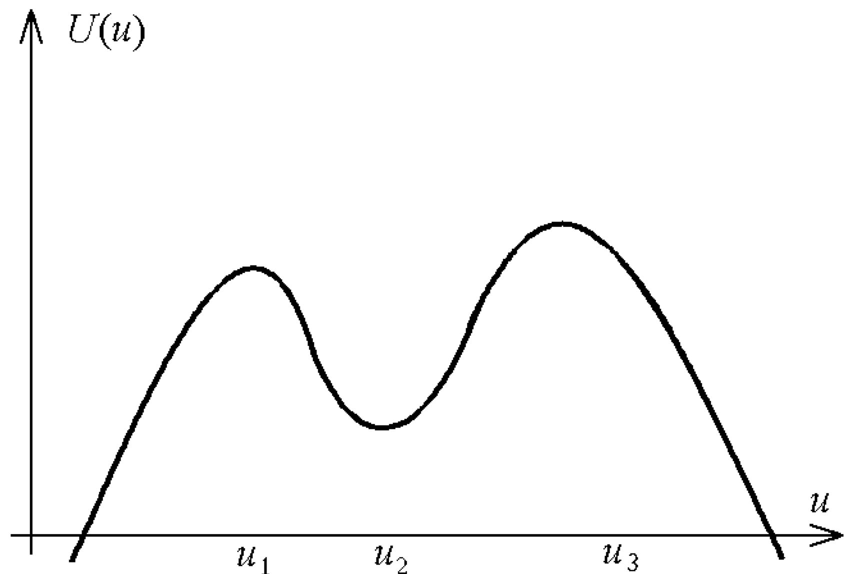


Рис. 8.5. К вычислению скорости волны переключения: график потенциала

Для того чтобы частица из состояния $u = u_3$ при $\xi = -\infty$ перешла в состояние $u = u_1$ при $\xi = \infty$, как это требуют граничные условия в (8.21), необходимо, чтобы энергия, высвобождающаяся при переходе от $u = u_3$ к $u = u_1$, равная $U_3 - U_1 = \int\limits_{u_1}^{u_3} f(z)dz \equiv A$, полностью истратилась на сопротивление трению. При заданном потенциале U существует единственное значение скорости волны c (трактуемое как вязкое трение в (8.23)), обеспечивающее выполнение этого условия, оно и дает выражение для скорости волны переключения.

Для анализа уравнения (8.20) запишем его в виде

$$\frac{\partial u}{\partial t} = -\frac{\delta F[u]}{\delta u}, \tag{8.24}$$

где

$$F = \int \left[-U(u) + \frac{1}{2}D(\nabla u)^2\right] dr, \tag{8.25}$$

а вариационная производная определена равенством

$$\frac{\delta F[u]}{\delta u}(r_0, t) = \lim_{\int \delta u dr \to 0} \frac{F[u + \delta u] - F[u]}{\int \delta u dr},$$

и вариация $\delta u(\cdot)$ локализована вблизи точки $r = r_0$. Тогда

$$\frac{dF}{dt} = \int \frac{\delta F[u]}{\delta u}(r, t)\frac{\partial u}{\partial t} dr,$$

или, с учетом (8.24),

$$\frac{dF}{dt} = -\int \left(\frac{\delta F[u]}{\delta u}\right)^2 dr, \tag{8.26}$$

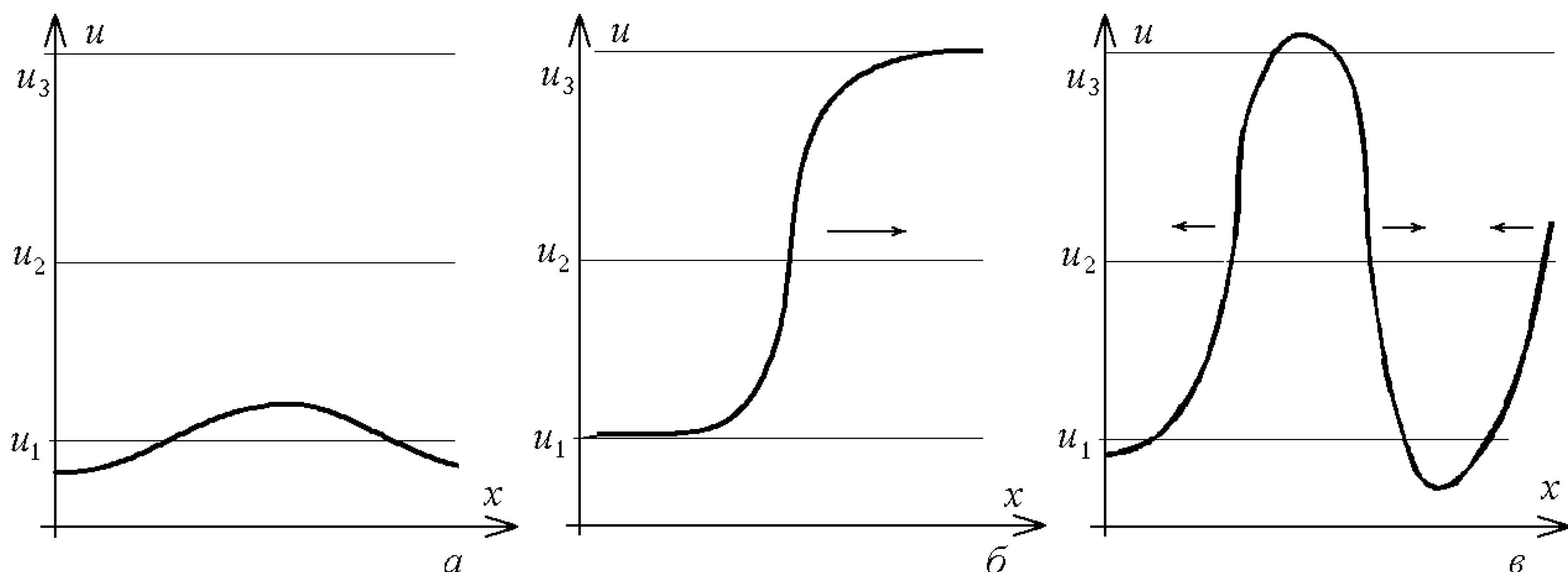


Рис. 8.6. Различные начальные условия в среде бистабильных элементов

т.е. эволюция системы сводится к тому, что функционал F не возрастает. Из определения F (8.25) следует, что локальными минимумами функционала F являются $u = u_1$ и $u = u_3$, т.е. эти решения соответствуют устойчивым состояниям системы, и к ним будут стремиться решения (8.20) при начальных профилях u, мало отличающихся от констант $u = u_1$ или $u = u_3$, рис. 8.6,*а*.

Начальный профиль, изображенный на рис. 8.6,*б*, неустойчив, так как при переходе к решениям $u = u_1$ или $u = u_3$ мы уменьшим значение функционала F. Профили общего вида, пример которого изображен на рис. 8.6,*в*, можно разбить на участки, на которых u "притягиваются" каждый к своему "ближайшему" состоянию — либо к $u = u_1$, либо к $u = u_3$, после чего в результате распространения волн переключения система перейдет к однородному по пространству устойчивому состоянию.

То из однородных состояний, в котором функционал F принимает большие значения, является метастабильным, из него возможен переход в основное устойчивое состояние. Такое поведение демонстрируют физические системы с фазовыми переходами первого рода. Состояния $u = u_1$ и $u = u_3$ называют фазами системы, метастабильная фаза устойчива к малым и неустойчива к большим возмущениям; при больших возмущениях система путем распространения волны переключения переходит к стабильному состоянию. Так ведет себя переохлажденная или перегретая жидкость, переключающаяся, соответственно, в твердую или газообразную фазы.

8.3.2. Возбудимые среды. Так называются среды, составленные из возбудимых элементов. Возбудимые, или мультивибраторные, элементы имеют единственное состояние покоя, устойчивое по отношению к слабым воздействиям, однако если внешнее воздействие на элемент превышает определенный порог, то он совершает заданную последовательность активных переходов, после чего вновь возвращается к основному невозбужденному состоянию. В моменты активных переходов элемент невосприимчив к внешним воздействиям.

Если цепочка составлена из возбудимых элементов, и два соседних элемента взаимодействуют тогда и только тогда, когда один из них находится в возбужденном состоянии, а другой — в невозбужденном, то возбуждение крайнего элемента приводит к постоянно действующей волновой активности. Если же воздействие на возбуждаемый элемент возможно лишь на ранней стадии активности возбуждающего элемента, то по цепочке распространяется волна возбуждения, после прохождения которой все элементы вновь оказываются в исходном стационарном состоянии, рис. 8.7. Этот пример, по сути, повторяет примеры реакции Белоусова–Жаботинского, распространения волн эпидемий, степных пожаров и т.п.

0	1	2	3	0	0	0	0	0	0	0	0
0	0	1	2	3	0	0	0	0	0	0	0
0	0	0	1	2	3	0	0	0	0	0	0
0	0	0	0	1	2	3	0	0	0	0	0
0	0	0	0	0	1	2	3	0	0	0	0

Рис. 8.7. Эволюция цепочки из возбудимых элементов

Для возбудимых элементов механизм, возвращающий систему к исходному состоянию, может быть описан следующим образом. Пусть режим горения в ячейке среды определяется не только температурой ячейки ϑ, но и концентрацией v некоторого вещества (ингибитора) так, что выделяемое тепло определяется заданной функцией $q = q(\vartheta, v)$. Чем выше концентрация ингибитора, тем меньше выделяется тепла. При заданном v распространение волны горения описывается уравнением

$$\frac{\partial\vartheta}{\partial t} = f(\vartheta, v) + k\frac{\partial^2\vartheta}{\partial x^2}, \qquad (8.27)$$

где

$$f(\vartheta, v) = -\gamma(\vartheta - \vartheta_{\text{env}}) + q(\vartheta, v)/c.$$

Как следует из анализа предыдущего примера, если

$$A(v) = \int\limits_{\vartheta_1}^{\vartheta_3} f(\vartheta, v)d\vartheta > 0,$$

то при распространении волны переключения из холодного состояния $\vartheta = \vartheta_1$ в состояние с высокой температурой $\vartheta = \vartheta_3$, см. рис. 8.2 и 8.5, и распространяющаяся волна есть волна загорания. Если же $A < 0$, то распространяется волна гашения. Но A зависит от концентрации ингибитора v, поэтому меняя v, можно управлять волной переключения, изменяя ее скорость и направление распространения. Так, при низких концентрациях v можно получить волну горения, а при высоких — волну гашения.

Если ингибитор выделяется в процессе горения и может уходить в окружающую среду, то его концентрацию можно описать уравнением

$$\frac{\partial v}{\partial t} = -\tau^{-1}(v - \overline{v}(\vartheta)), \qquad (8.28)$$

где функция $\overline{v}(\vartheta)$ — равновесная концентрация ингибитора, монотонно возрастающая с ростом температуры ϑ. Будем считать, что характерное время установления концентрации ингибитора τ велико по сравнению с длительностью процесса перехода от холодного к горячему состоянию, т.е. — по сравнению с длительностью вспышки загорания, и диффузия ингибитора мала, так, что ею можно пренебречь. Тогда, совместно с уравнением (8.27), уравнение (8.28) дает модель распространения импульса в среде, рис. 8.8. Фронт и спад импульса могут рассматриваться как две волны — загорания и гашения, движущиеся друг за другом с одинаковой скоростью. Гашение осуществляется при достаточно большой концентрации ингибитора. Такими уравнениями описываются явления распространение волн в реакции Белоусова–Жаботинского, волн степных пожаров, волн эпидемий и т.п.

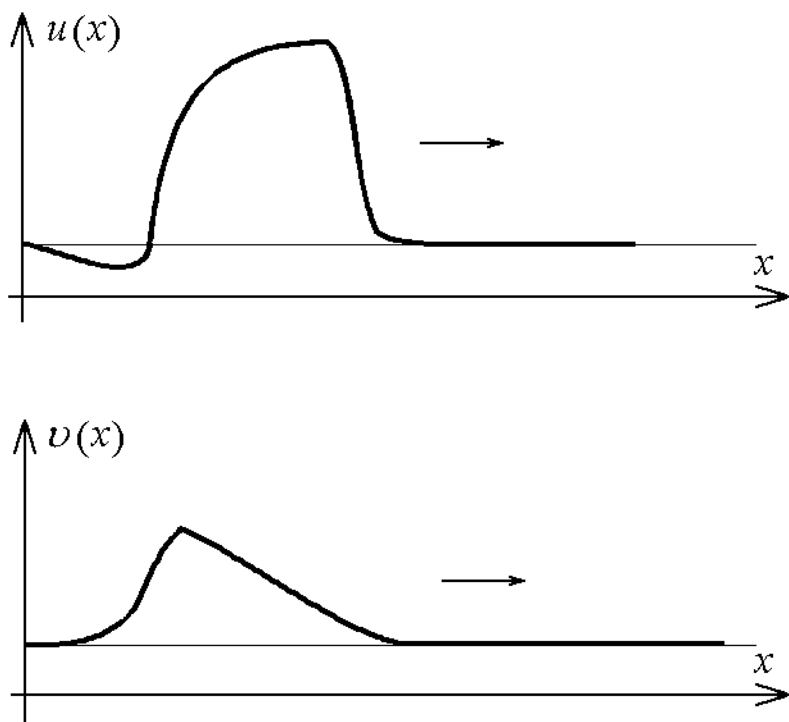


Рис. 8.8. Распространение волны в возбудимых средах

Итак, класс возбудимых сред описывается системой дифференциальных уравнений вида

$$\begin{aligned} \frac{\partial u}{\partial t} &= f(u,v) + D\Delta u, \\ \varepsilon^{-1}\,\frac{\partial v}{\partial t} &= -v + \overline{v}(u), \end{aligned} \qquad (8.29)$$

где $\varepsilon \ll 1$, а $\overline{v}(\cdot)$ — монотонно возрастающая функция. Выбором параметров системы уравнений (8.29) можно добиться того, чтобы у нее было единственное устойчивое однородное по пространству состояние. Это произойдет, когда существует единственное решение системы алгебраических уравнений

$$\begin{aligned} f(u,v) &= 0, \\ v &= \overline{v}(u), \end{aligned}$$

Для исследования уравнений (8.29) можно воспользоваться приближенными представлениями о процессе, считая, что концентрация v в

процессе возгорания и в процессе гашения не меняется, тем самым в (8.29) можно рассматривать волны переключения, определяемые первым уравнением (8.29), и воспользоваться идеями, изложенными при анализе бистабильных сред.

В активных возбудимых средах возможно существование и периодических решений вида $u = u(\eta)$, $v = v(\eta)$, $\eta = kx - \omega t$. Это решение удовлетворяет системе обыкновенных дифференциальных уравнений

$$-\omega u' = f(u,v) + Dk^2 u'',$$
$$-\tfrac{\omega}{\varepsilon} v' = \overline{v}(u) - v,$$

и соответствует предельному циклу этой системы. Как правило, такие волны являются диспергирующими со своими дисперсионными соотношениями $\omega = \Omega(k,\varepsilon)$, см. раздел ().

8.3.3. Автоколебательные среды. Так называются среды, составленные из совокупности соединенных между собой автоколебательных элементов, которые характеризуются тем, что постоянно совершают циклические переходы через некоторую определенную последовательность состояний, а внешние воздействия могут лишь ускорить или замедлить эти процессы.

Пусть отдельно взятый элемент среды совершает автоколебания, которые характеризуются своей формой, амплитудой, частотой и начальной фазой. При этом форма, частота и амплитуда определяются свойствами элемента, а начальная фаза может быть произвольной. После возмущения элемента он возвращается к колебаниям прежней формы, амплитуды и частоты, но фаза может стать иной. Будем рассматривать описание среды только в терминах фазы.

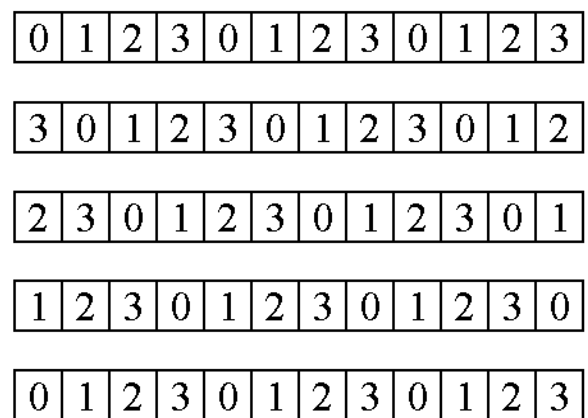


Рис. 8.9. Эволюция цепочки из автоколебательных элементов

Если составить цепочку из автоколебательных элементов, то в ней возбуждаются фазовые волны, рис. 8.9.

Пусть в среде, составленной из автоколебательных элементов, взаимодействие между соседними элементами пропорционально рассогласованию между ними по форме, частоте, амплитуде или фазе. При этом существует режим синхронных колебаний, когда $\varphi(r,t) = \mathrm{const}$. Если фаза меняется вдоль среды плавно и характеризуется достаточно большим масштабом L, то распределение будет достаточно медленно меняться и по времени, так как чем больше характерный масштаб L, тем меньше рассогласование по фазе соседних элементов.

Для активных автоколебательных сред характерные времена релаксаций возмущений амплитуды и фазы различаются, первые до-

статочно малы, а вторые — велики, поэтому динамика фазовых изменений может учитываться отдельным уравнением.

В однородной активной среде изменение фазы $\partial\varphi/\partial t$ не может зависеть от направления градиента фазы, поэтому эволюционный оператор фазы может содержать в разложении по $\nabla\varphi$ только его четные степени. Учет возможных пространственных изменений приводит к уравнению

$$\frac{\partial\varphi}{\partial t} = a(\nabla\varphi)^2 + b\Delta\varphi, \qquad (8.30)$$

где a и b — некоторые коэффициенты.

В частности, (8.30) имеет решение вида

$$\varphi = -kr + ak^2 t.$$

Таким образом, частоты колебаний среды

$$\omega = \omega_0 + ak^2, \qquad (8.31)$$

в зависимости от знака a в (8.30) увеличиваются или уменьшаются при взаимодействии, и (8.31) дает дисперсионное соотношение для фазовых волн.

Несколько более сложными являются среды, составленные из ячеек, расположенных на плоскости или в пространстве. К рассмотрению таких сред мы вернемся в разделе, посвященном клеточным автоматам.

8.4. Нелинейные волны. Солитоны

В этом разделе рассматриваются непрерывные модели сред, в которых возникают структуры в виде распространяющихся волн.

Строгого математического определения понятия волны, описывающего все волновые явления, происходящие в природе, по-видимому, не существует. Можно дать ряд частных определений, однако можно и ограничиться интуитивным представлением о волнах как о любом сигнале, передающемся от одной части среды к другой. При этом природа сигнала не важна — это может быть любое возмущение среды, например, максимум какой-либо величины (давления, плотности), или резкое ее изменение, главное — чтобы была возможность четко фиксировать местонахождение этого возмущения в любой момент времени.

8.4.1. Гиперболические и диспергирующие волны. Выделяют два основных класса волн. Первый класс описывается уравнениями гиперболического типа и носит название гиперболических волн.

Ко второму классу относятся все диспергирующие волны, т.е. волны, скорость которых зависит от частоты. Заметим, что эти классы пересекаются, и существуют исключения, не принадлежащие ни к одному из названных типов.

Описание гиперболических волн основано на уравнении Д'Аламбера

$$\varphi_{tt} = c_0^2 \nabla^2 \varphi, \tag{8.32}$$

или на более простом —

$$\varphi_t = c_0^2 \varphi_x. \tag{8.33}$$

Диспергирующие волны, в отличие от гиперболических, определяются не столько уравнениями, сколько свойствами их решений. Линейная диспергирующая система — это любая система, имеющая решения вида

$$\varphi = a \cos(\kappa x - \omega t), \tag{8.34}$$

где $\omega = \omega(\kappa)$ — заданная функция волнового числа κ. Зафиксируем фазу волны: $\kappa x - \omega t = \mathrm{const}$, откуда получим уравнение движения точки постоянной фазы в виде

$$x = \frac{\mathrm{const} + \omega t}{\kappa},$$

а значит, скорость ее движения (фазовая скорость) равна $\omega(\kappa)/\kappa$. Волны называются диспергирующими, если эти скорости не постоянны, а зависят от κ. Интерес к диспергирующим волнам связан с тем, что решение рассматривается как сумма нескольких слагаемых вида (8.34), или как интеграл Фурье. Если фазовые скорости зависят от κ, т.е. $\omega \neq c_0 \kappa$, где $c_0 = \mathrm{const}$, то моды с различными волновыми числами κ будут распространяться с разными скоростями, и волна будет диспергировать, т.е. расползаться. (В случае нелинейных волн это расползание может компенсироваться нелинейностью — тогда форма волны сохраняется. Такие волны называются солитонами).

Линейное уравнение вида (8.32) или (8.33) тоже дает решение вида (8.34), однако для этих решений $\omega = c_0 \kappa$, т.е. в таких волнах дисперсия отсутствует. В то же время существуют гиперболические уравнения, дающие решения с нетривиальными дисперсионными соотношениями. Такой пример дает уравнение Клейна–Гордона:

$$\varphi_{tt} - \varphi_{xx} + \varphi = 0, \tag{8.35}$$

оно является гиперболическим и имеет дисперсионное соотношение

$$\omega^2 = \kappa^2 + 1. \tag{8.36}$$

В общем виде диспергирующие волны характеризуются дисперсионным соотношением вида

$$\omega = W(\kappa). \tag{8.37}$$

В линейном случае оно может быть получено методом Фурье. Для этого берется пробное решение в виде

$$\varphi(x,t) = Ae^{i\kappa t - i\omega t} \tag{8.38}$$

и подставляется в исходное уравнение (здесь κ — вещественное, а ω может быть и комплексным). Дифференциальное уравнение приводит при этом к соотношению

$$\widetilde{W}(\omega,\kappa)Ae^{i\kappa t - i\omega t} = 0,$$

откуда и из требования нетривиальности решения получается дисперсионное соотношение (8.37).

Помимо уравнения Клейна–Гордона (8.35), диспергирующие решения дают линейное уравнение колебаний балки:

$$\varphi_{tt} + \gamma^2\varphi_{xxxx} = 0,$$

для него дисперсионное соотношение дается формулой

$$\omega = \pm\gamma\kappa^2,$$

линейное уравнение Кортевега–де Фриза (КДВ):

$$\varphi_t + c_0\varphi_x + \nu\varphi_{xxx} = 0,$$

для него описанная процедура приводит к дисперсионному соотношению

$$\omega = c_0\kappa - \nu\kappa^3,$$

уравнение Буссинеска

$$\varphi_{tt} - \alpha^2\varphi_{xx} - \beta^2\varphi_{xxtt} = 0$$

с дисперсионным соотношением

$$\omega = \pm\alpha\kappa(1 + \beta^2\kappa^2)^{-1/2}.$$

Два последних уравнения возникают как приближение длинных волн на воде.

Нелинейные уравнения, описывающие диспергирующие волны, уже не имеют решений вида (8.38), и вместо экспоненты в (8.38)

стоит, как правило, другая периодическая функция, возможно, и с бесконечным периодом. Дисперсионные соотношения в нелинейных случаях содержат, как правило, и амплитуду волны. Интересным свойством нелинейных уравнений, описывающих волны, является существование так называемых уединенных волн — солитонов, в которых нелинейность компенсирует дисперсию, что приводит к сохранению неизменной формы волны.

Одним из наиболее значительных успехов теории нелинейных волн является открытие точных решений для некоторых простых канонических уравнений — это, в основном, три следующих уравнения.

1. Уравнение Кортевега–де Фриза (КДВ):

$$\eta_t + \sigma\eta\eta_x + \eta_{xxx} = 0, \tag{8.39}$$

где $\sigma = \mathrm{const}$.

2. Кубическое уравнение Шрёдингера:

$$iu_t + u_{xx} + \nu u|u|^2 = 0. \tag{8.40}$$

3. Уравнение Sin–Гордона:

$$\varphi_{tt} - \varphi_{xx} + \sin\varphi = 0. \tag{8.41}$$

Для них были построены явные решения, описывающие взаимодействия произвольного числа уединенных волн, и предсказывающие точное число уединенных волн, к которому в конце концов сводится любое финитное возмущение. Помимо того, что эти уравнения часто возникают в приложениях, они интересны тем, что объединяют простейшие типы дисперсии с простейшими типами нелинейности. Так, уравнение КДВ (8.39) объединяет линейную дисперсию

$$\omega = \kappa^3 \tag{8.42}$$

с простейшим нелинейным оператором переноса $\eta_t = \sigma\eta\eta_x$. Уравнение Шрёдингера (8.40) объединяет дисперсию

$$\omega = \kappa^2 \tag{8.43}$$

с простой кубической нелинейностью. Как квадратичная дисперсия (8.42), так и кубическая (8.43) могут рассматриваться как начало разложения более общего дисперсионного соотношения (8.37) в ряд Тейлора по степеням κ, причем (8.43) относится к случаю четных степеней, а (8.42) — к случаю нечетных. Тем самым эти уравнения можно рассматривать как длинноволновые приближения для более общего случая (8.37).

Линейное дисперсионное соотношение вида (8.36) в уравнении Sin–Гордона объединено с периодичностью, которая возникает, например, при нелинейном эффекте Джозефсона (там $\sin\varphi$ описывает ток между двумя сверхпроводниками), при описании дислокаций в кристалле, где появление $\sin\varphi$ связано с периодичностью кристаллической решетки, и в других приложениях.

8.4.2. Солитоны. История солитонов началась с наблюдения Джона Скотта Рассела в 1834 г. уединенной волны в канале, соединяющем два побережья Шотландии. Эта волна возникла при резкой остановке баржи и двигалась по каналу не затухая со скоростью пешехода, что дало возможность наблюдать ее на протяжении нескольких километров, до тех пор, пока канал не стал дробиться на узкие рукава.

Долгое время нахождение солитоноподобных решений считалось искусством, однако в 60–70 годах XX века появились некоторые общие подходы к решению этих задач, в первую очередь это метод обратной задачи, когда решение нелинейной проблемы представлялось в виде потенциала в линейной задаче рассеяния, и находилось стандартной процедурой из данных рассеяния на бесконечности. Другой подход состоял в нелинейных преобразованиях уравнений, сводящих задачу к известной — типа (8.39)–(8.41); они носят название преобразований Бэклунда–Ли.

Взаимодействие солитонов. Исследования солитонов начались с удачной замены переменных, которая позволила найти точное решение уравнения КДВ в виде уединенной волны (солитона):

$$\eta(x,t) = \frac{3}{\sigma}\alpha^2 \operatorname{sech}\left(\frac{\vartheta-\vartheta_0}{2}\right), \quad \vartheta = \alpha x - \alpha^3 t. \tag{8.44}$$

Этот подход можно использовать для описания взаимодействия нескольких солитонов, представленных до взаимодействия суммой нескольких функций вида (8.44) с различающимися параметрами. Общий подход к решению уравнений типа (8.39)–(8.41) связан с методом обратной задачи рассеяния, идею которого мы здесь рассмотрим, оставляя в стороне подробности аналитических преобразований.

Исследуем сначала взаимодействие солитонов. Рассмотрим для примера уравнение КДВ (8.39) и сделаем замену переменной

$$\sigma\eta = 12(\ln F)_{xx}. \tag{8.45}$$

Из общих соображений ясно, что такая замена приведет к появлению третьих и четвертых степеней в уравнении (8.39), однако в окончательном выражении эти степени взаимно уничтожатся, и в результате получится квадратичное уравнение

$$F(F_t + F_{xxx})_x - F_x(F_t + F_{xxx}) + 3(F_{xx}^2 - F_x F_{xxx}) = 0. \tag{8.46}$$

Идея, на основе которой была произведена замена (8.45), состоит в том, что решение (8.44), описывающее уединенную волну, соответствует функции

$$F = 1 + e^{-(\vartheta - \vartheta_0)} = 1 + e^{-\alpha(x-s)+\alpha^3 t}, \quad s = \vartheta_0/\alpha, \tag{8.47}$$

т.е. преобразование (8.45) переводит функцию, описывающую уединенную волну, в простую экспоненту. Показатель экспоненты выбран так, что функция вида (8.47) удовлетворяет уравнению

$$\left(\frac{\partial}{\partial t} + \frac{\partial^3}{\partial x^3}\right) F = 0$$

и тем самым (8.47) удовлетворяет и (8.46) для любых α и s, так как третий член в (8.46) обращается в нуль благодаря своей однородности по производным.

Из-за нелинейности уравнения (8.46) сумма экспонент вида (8.47) не является его решением, поэтому воспользуемся приемом, обычным для теории взаимодействий, и будем искать решение в виде

$$F = 1 + F^{(1)} + F^{(2)} + \ldots,$$

где функции $F^{(i)}$, $i = 1, 2, \ldots$ находятся из цепочки уравнений

$$\begin{gathered} \left\{F_t^{(1)} + F_{xxx}^{(1)}\right\}_x = 0, \\ \left\{F_t^{(2)} + F_{xxx}^{(2)}\right\}_x = -3\left\{(F_{xx}^{(1)})^2 - F_x^{(1)} F_{xxx}^{(1)}\right\}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{gathered} \tag{8.48}$$

Если в качестве $F^{(1)}$ взять сумму двух экспонент $f_1 + f_2$ вида (8.47) с разными коэффициентами α_i и s_i, $i = 1, 2$, то в уравнениях для $F^{(i)}$, $i = 3, 4, \ldots$, правые части обращаются в нуль, и решая уравнение для $F^{(2)}$, получим точное решение (8.46) в виде

$$F = 1 + f_1 + f_2 + \frac{(\alpha_2 - \alpha_1)^2}{(\alpha_2 + \alpha_1)^2} f_1 f_2, \tag{8.49}$$

здесь последнее слагаемое в (8.49) описывает взаимодействие двух уединенных волн.

К аналогичному результату прийдем, если начнем с функции

$$F^{(1)} = \sum_{j=1}^{N} f_j,$$

в этой ситуации последовательность (8.48) обрывается на N-м шаге, при этом в формуле типа (8.49) последнее слагаемое будет пропорционально произведению $f_1 f_2 \dots f_N$. Это слагаемое описывает взаимодействие N уединенных волн.

Исследуем решение (8.49), положив для определенности, что выполнено неравенство $\alpha_1 > \alpha_2 > 0$. Анализ формулы (8.49) показывает, что уединенная волна с параметрами α_2, s_2 имеет меньшую амплитуду и скорость, чем волна с параметрами α_1, s_1. При $t \to -\infty$ решение (8.39) представимо в виде двух уединенных волн

$$\frac{\sigma}{12}\eta = \frac{\alpha_1^2 f_1}{(1+f_1)^2} + \frac{\alpha_2^2 f_2}{(1+f_2)^2} \tag{8.50}$$

и соответствует случаю, когда большая уединенная волна догоняет меньшую.

При $t \to \infty$ решение тоже представимо в виде суммы двух уединенных волн, но теперь волна с большей амплитудой обогнала волну с меньшей амплитудой. Тем самым, уединенные волны выходят из области взаимодействия без изменения своей формы и с первоначальными параметрами α_1 и α_2 так же, как это было бы в линейном случае. Единственное напоминание о нелинейном взаимодействии заключается в том, что волна большей амплитуды оказывается запаздывающей (по сравнению с невзаимодействующим распространением), а с меньшей — опережающей. Само взаимодействие происходит в области, где f_1 и f_2 имеют значения порядка единицы, см. (8.50), и точка взаимодействия легко может быть вычислена:

$$t_* = -\frac{s_2 - s_1}{\alpha_2^2 - \alpha_1^2}, \quad x = \frac{\alpha_2^2 s_1 - \alpha_1^2 s_2}{\alpha_2^2 - \alpha_1^2}.$$

О методе обратной задачи рассеяния. Обратимся теперь к общему методу поиска решения уравнений типа (8.39)–(8.41) в виде солитонов, получившему название метода обратной задачи рассеяния. Ограничимся при этом лишь изложением общей идеи метода. Положим в (8.39) $u = -(\sigma/6)\eta$, чтобы согласовать уравнение с оригинальными работами, тогда уравнение КДВ примет "привычный" вид

$$u_t - 6uu_x + u_{xxx} = 0.$$

Запишем подстановку

$$u \sim \frac{\psi_{xx}}{\psi} + \lambda,$$

и заметим, что это равенство можно записать как задачу на собственные значения оператора $\left(\partial^2/\partial x^2 - u\right)$, т.е. как приведенное уравнение Шрёдингера

$$\psi_{xx} + (\lambda - u)\psi = 0,$$

которому можно придать смысл задачи рассеяния на потенциале u; такой подход к этой задачи связан с тем, что существует достаточно хорошо развитый аппарат решения задач рассеяния, который по поведению функции ψ на бесконечности позволяет высказываться о потенциале u.

Метод решения опирается на утверждение, что рассеивающий потенциал u можно построить, зная коэффициенты отражения для волн, пришедших из $+\infty$, и располагая некоторой информацией о точечном спектре. Это так называемая обратная задача рассеяния, состоящая в определении характеристик рассеивателя по его отражающим свойствам. В нашем же случае необходимые данные о решении получаются не из эксперимента, а из уравнений.

Похожие методы были разработаны и для уравнений Sin–Гордона, и для нелинейного уравнения Шрёдингера.

8.5. Самоорганизация в химической кинетике

В этом разделе рассмотрим подробнее математические модели уже упоминавшихся выше процессов химической кинетики, в которых интересуются временной, а иногда и пространственной зависимостью концентраций реагирующих веществ. Дело в том, что не все химические процессы идут так, что, определив начальные реагенты, можно определенно сказать, что получится в результате к концу реакции. Обнаружены химические процессы, в которых концентрации веществ претерпевают сложные изменения, а частности, автоколебания, так что “конца реакции” в обычном понимании этого слова может и не быть. Исторически первой среди автоколебательных химических явлений была реакция Белоусова–Жаботинского.

8.5.1. Реакция Белоусова–Жаботинского. Это один из наиболее подробно изученных примеров колебательных химических процессов. Реакция Белоусова–Жаботинскогосостоит в окислении лимонной кислоты (в более поздних экспериментах замененной на другую органическую кислоту — малоновую) броматом калия, реакция протекает в кислой среде в присутствии ионов церия Ce^{3+}, которые являются и катализатором, и окрашивающим индикатором.

Интерес к реакции Белоусова–Жаботинского состоит в том, что гомогенный раствор с большой точностью периодически изменяет свою окраску (в первоначальной реакции, выполненной Белоусовым, — с бледно-желтой на бесцветную, в более поздних вариантах удалось добиться и более ярких цветов). Такой характер реакции позволяет называть ее химическими часами.

По мере продолжения исследований были найдены и другие ин-

тересные свойства этой реакции, при этом несколько изменились и ее компоненты. Если реакцию Белоусова–Жаботинского проводить в тонком слое раствора в чашке Петри, то на поверхности раствора начинает проявляться и распространяться геометрическая картина, состоящая из окрашенных узких концентрических колец с увеличивающимся со временем радиусом — происходит распространение концентрических волн. Волны имеют скорость около 1 мм/с, и распространяются из одного или нескольких центров — пейсмейкеров.

Если на пути волны в чашке Петри поставить преграду (достаточно простой булавки), то волна разбивается на две спиральных волны, закручивающиеся в разные стороны. При этом на локальном участке, на котором вращается одна спиральная волна, происходит так называемый разрыв симметрии, появляется выделенное направление вращения, тем самым увеличивается самоорганизация.

Механизм распространения волн концентрации в реакции Белоусова–Жаботинского аналогичен механизму, порождающему волны эпидемии. Пусть некоторый вирус гриппа вызывает заболевание, длящееся в течение нескольких дней, после чего выздоровевший обладает кратковременным иммунитетом. Вирус распространяется и заражает людей в некотором районе, а люди, перемещаясь, переносят заболевание в другие районы. Так происходит движение фронта волны гриппа. Выздоровевшие люди, обладающие иммунитетом, потом в течение относительно долгого времени не болеют, но если источник вируса действует постоянно, то могут возникать концентрические волны распространения гриппа, похожие по сути на волны концентраций веществ при реакции Белоусова–Жаботинского в чашке Петри. Такой же пример дают степные пожары: для них нужно сочетание огня и сухой травы: после прохождения огня нужно некоторое время для того, чтобы трава выросла и высохла, после чего может вновь возникнуть волна пожара.

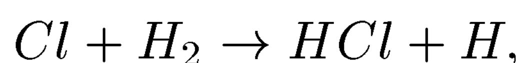
8.5.2. Математическая модель реакций химической кинетики. Обратимся к математической модели, описывающей временную и пространственную зависимости концентрации компонент реакции Белоусова–Жаботинского. Для этого следует указать механизм реакции, т.е. указать, какие атомы и молекулы взаимодействуют между собой и какие химические вещества получаются в результате. Такие взаимодействия называются элементарными реакциями. Было найдено, что суммарная реакция Белоусова–Жаботинского состоит из более, чем 20 элементарных, многие из которых до конца еще не исследованы. Построим математическую модель на основе лишь небольшого числа самых характерных элементарных реакций, ответственных за колебательный процесс.

Для того чтобы по химическому механизму написать уравнения для концентраций веществ, участвующих в реакции, были найдены

определенные правила. Пусть, например, реакция является мономолекулярной, т.е. состоит в том, что молекула какого-либо вещества A путем перегруппировки атомов или путем распада превращается в один или более продуктов реакции. Тогда скорость такой реакции уменьшается пропорционально концентрации исходного вещества, и уравнение для его концентрации выглядит следующим образом:

$$\frac{dA}{dt} = -kA.$$

Коэффициент k называется скоростью реакции и определяется экспериментально. Если реакция протекает в результате соударения двух молекул или атомов, например,

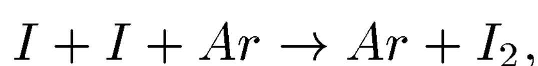


$$Cl + H_2 \to HCl + H,$$

то реакция называется бимолекулярной, и уравнение для концентрации, например, хлора записывается в виде

$$\frac{d[Cl]}{dt} = -k[Cl][H_2],$$

здесь $[B]$ обозначает концентрацию вещества с химической формулой B. Если в акте соударения для химических превращений необходимо участие трех молекул или атомов, то говорят о тримолекулярных реакциях. К такому типу реакций относится образование молекулы йода I_2 из двух его атомов в присутствии инертного газа аргона Ar:



$$I + I + Ar \to Ar + I_2,$$

концентрация атомарного йода при этом описывается уравнением

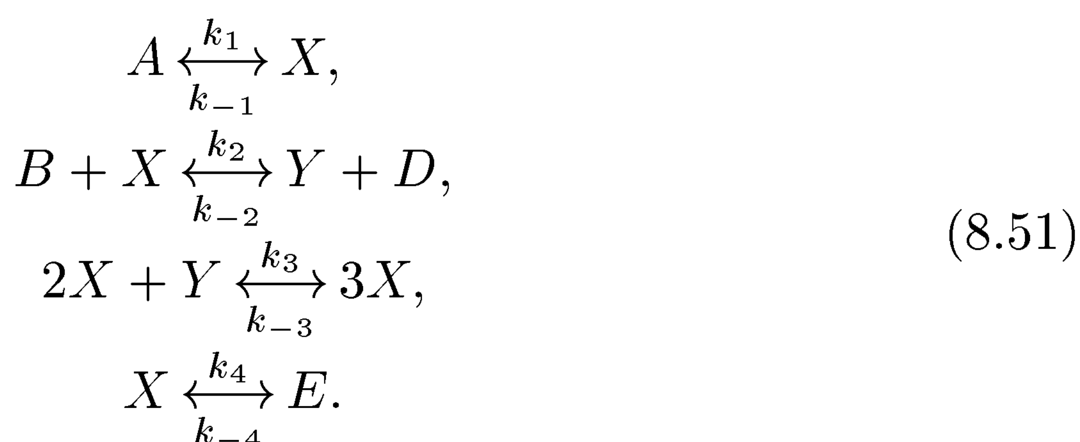
$$\frac{d[I]}{dt} = -k[I]^2[Ar].$$

Такие реакции достаточно редки и медленны.

8.5.3. Брюсселятор. Одной из наиболее простых моделей, описывающих колебательные химические реакции, является созданная в Брюсселе в 1968 г. И. Пригожиным и Р. Лефевром модель, названная по месту создания брюсселятором. При создании этой модели исходили из достаточно простых принципов. Например, для создания колебательной реакции по меньшей мере необходимо два вещества с изменяющейся концентрацией (в примере со степью — это огонь и сухая трава, в примере с эпидемией гриппа — концентрации вируса и способных им заразиться). Поэтому уравнения брюсселятора составлены относительно концентрации двух веществ, участвующих

в реакции — $X(t)$ и $Y(t)$. Концентрация первого растет до тех пор, пока оно не вступает в реакцию со вторым, что затормозит скорость ее роста, и даже приведет к уменьшению концентрации X, зато увеличит концентрацию Y. Если большая концентрация Y приведет к росту концентрации X, то реакция начнет носить колебательный характер.

Можно показать, что мономолекулярные и бимолекулярные реакции при наличии двух веществ с изменяющейся концентрацией не могут привести к колебаниям их концентраций, и необходимо включение в модель тримолекулярных и автокаталитических реакций. Исходя из этих соображений была написана следующая система реакций, лежащая в основе описываемой модели:



$$\begin{gathered} A \underset{k_{-1}}{\overset{k_1}{\longleftrightarrow}} X, \\ B + X \underset{k_{-2}}{\overset{k_2}{\longleftrightarrow}} Y + D, \\ 2X + Y \underset{k_{-3}}{\overset{k_3}{\longleftrightarrow}} 3X, \\ X \underset{k_{-4}}{\overset{k_4}{\longleftrightarrow}} E. \end{gathered} \tag{8.51}$$

Здесь первая строка (8.51) означает, что вещество A со скоростью реакции k_1 превращается в вещество X, и идет обратная реакция со скоростью k_{-1}; аналогично объясняются и остальные записи (8.51).

Распространенность модели брюсселятора (8.51) объясняется тем, что она на качественном уровне описывает все процессы, наблюдаемые при реакции Белоусова–Жаботинского, и кроме того, благодаря ей были открыты и иные возможности протекания химических реакций, которые пока еще не были осуществлены на практике.

Пространственно однородная модель. Предполагается, что реакция в брюсселяторе протекает в открытой системе — в реакторе, в котором путем регулирования входных потоков J_A и J_B и выходных потоков J_D и J_E концентрации веществ A, B, D и E поддерживаются постоянными и вдали от положения динамического равновесия (рис. 8.10). Эти концентрации рассматриваются как управляющие параметры, и в дифференциальные эволюционные уравнения входят лишь концентрации $X(t)$ и $Y(t)$. Если в реакторе смесь веществ постоянно перемешивается, то он относится к типу пространственно однородных систем. В таком реакторе может существовать равновесное состояние, т.е. такое, при котором концентрации веществ не меняются, так как их изменение в прямой реакции уравновешивается

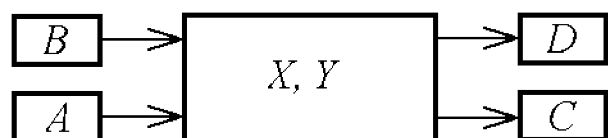


Рис. 8.10. Схема брюсселятора

обратной. Воспользовавшись (8.51), уравнения для концентрации при равновесном состоянии получим в виде

$$k_{-1}X = k_1 A, \quad k_{-4}E = k_4 X,$$
$$k_2 BX = k_{-2}YD, \quad k_3 X^2 = k_{-3}X^3,$$

что приведет к следующим значениям равновесных концентраций X^{eq} и Y^{eq} веществ X и Y:

$$X^{eq} = \frac{k_1}{k_{-1}}A, \quad Y^{eq} = \frac{k_{-3}k_1}{k_3 k_{-1}}A,$$

при этом концентрации веществ A, B, D и E остаются постоянными и должны быть связаны соотношением

$$\frac{k_1}{k_{-1}}A = \frac{k_{-4}}{k_4}E, \quad \frac{k_2}{k_{-2}}B = \frac{k_{-3}}{k_3}D.$$

Пользуясь управляющими переменными, можно создать сильно неравновесное состояние, т.е. состояние, находящееся вдали от того, когда все концентрации уравновешиваются прямыми и обратными реакциями. Рассмотрим динамику концентраций в этом состоянии, для чего для простоты описания введем новые переменные, в которых скорости прямых реакций равны единице, а обратные реакции полностью подавлены (для этого в уравнениях коэффициенты с отрицательными индексами полагаются равными нулю). В этом случае эволюционные уравнения для концентраций X и Y запишутся в виде системы дифференциальных уравнений

$$\begin{aligned} \frac{dX}{dt} &= A - (B+1)X + X^2Y, \\ \frac{dY}{dt} &= BX - X^2Y. \end{aligned} \tag{8.52}$$

Пространственно неоднородная модель. Если принудительного перемешивания не происходит, то концентрация $X(t)$ и $Y(t)$ изменяется не только из-за химических превращений, но и благодаря процессу диффузии. Для учета этого в уравнения (8.52) следует добавить диффузионные члены; в результате получим пространственно неоднородную модель брюсселятора

$$\begin{aligned} \frac{\partial X}{\partial t} &= A - (B+1)X + X^2Y + D_X\frac{\partial^2 X}{\partial r^2}, \\ \frac{\partial Y}{\partial t} &= BX - X^2Y + D_Y\frac{\partial^2 Y}{\partial r^2}, \quad 0 < r < l; \end{aligned} \tag{8.53}$$

здесь считается, что реакция происходит в длинной тонкой трубке длины l, и концентрации $X(\mathbf{r}, t)$ и $Y(\mathbf{r}, t)$ однородны в сечении трубки.

Уравнение (8.53) описывает процесс типа “реакция — диффузия” и должно быть дополнено граничными условиями. Так как потоки веществ X и Y на границах трубки отсутствуют, т.е. нет утечки продуктов реакции, то естественным условием на границе является второе краевое условие:

$$\frac{\partial X}{\partial t}\bigg|_{r=0} = \frac{\partial Y}{\partial t}\bigg|_{r=0} = \frac{\partial X}{\partial t}\bigg|_{r=l} = \frac{\partial Y}{\partial t}\bigg|_{r=l} = 0. \tag{8.54}$$

8.5.4. Анализ математической модели брюселятора. Вернемся к пространственно однородной системе уравнений (8.52) и найдем ее стационарную точку. Она дается соотношениями

$$X^s = A, \quad Y^s = \frac{B}{A}. \tag{8.55}$$

Линеаризация системы (8.52) в окрестности точки (X^s, Y^s) приводит нас к следующей линейной системе обыкновенных дифференциальных уравнений относительно возмущений стационарного состояния $x(t) = X(t) - X^s$, $y(t) = Y(t) - Y^s$:

$$\begin{aligned} \frac{dx}{dt} &= -(B+1)x + 2Y^sX^sx + yX^{s2}, \\ \frac{dy}{dt} &= Bx - 2xY^sX^s - yX^{s2}, \end{aligned} \tag{8.56}$$

ее решение дается выражениями

$$x = C_1e^{\omega_1 t} + C_2e^{\omega_2 t},$$

$$y = C_1K_1e^{\omega_1 t} + C_2K_2e^{\omega_2 t},$$

где $\omega_{1,2}$ — корни квадратного уравнения

$$\omega^2 + (A^2 - B + 1)\omega + A^2 = 0,$$

т.е.

$$\omega_{1,2} = \frac{T \pm \sqrt{T^2 - 4\Delta}}{2}, \tag{8.57}$$

здесь введены обозначения $T = -A^2 + B - 1$, $\Delta = A^2$, а коэффициенты $K_{1,2}$ — корни квадратного уравнения

$$A^2K^2 + (B - 1 + A^2)K + B = 0.$$

Из соотношений (8.57) ясно, что корни являются действительными при $T^2 - 4\Delta \geq 0$, причем оба корня имеют один и тот же знак.

Если при этом $T > 0$, то оба корня положительны, что приводит к линеаризованным решениям (8.56) системы (8.52) в виде растущих экспонент и свидетельствует о неустойчивости решения.

Если $T < 0$, то оба корня (8.57) отрицательны, и такое стационарное состояние устойчиво.

Если $T^2 - 4\Delta = 0$, то $\omega_{1,2} = T/2$, и при $T > 0$ это неустойчивый узел, а при $T < 0$ — устойчивый узел.

Если $T^2 - 4\Delta < 0$, то частоты комплексные, что приводит к колебательному характеру решений.

В этом случае при $T > 0$ решение (8.52) в фазовой плоскости изображается расходящимися спиралями в окрестности стационарной точки, а при $T < 0$ — сходящимися к стационарной точке. При $T = 0$ фазовые траектории — замкнутые линии.

Итак, мы можем теперь описать бифуркационную диаграмму брюсселятора. Пусть, для простоты, A фиксировано, а B изменяющийся управляющий параметр. Тогда

— если $B < (A-1)^2$, то стационарная особая точка системы (8.52) является устойчивым узлом, частоты (8.57) — действительны;

— если $(A-1)^2 < B < (A+1)^2$, то стационарная особая точка системы (8.52) является устойчивым фокусом, частоты (8.57) — комплексные;

— если $B = (A+1)^2$, то стационарная особая точка системы (8.52) является центром, частоты (8.57) чисто мнимые, и в окрестности особой точки существуют колебания;

— при $B > (A+1)^2$, то стационарная особая точка системы (8.52) — неустойчивый фокус.

Для системы с диффузией следует найти стационарное решение системы (8.53) в виде их разложения по собственным функциям оператора диффузии, т.е. двукратного дифференцирования по r, удовлетворяющих краевым условиям (8.54). Решение (8.53) представляется в виде множества мод, каждая из которых демонстрирует поведение, похожее качественно на поведение пространственно однородной системы, однако возникает и новый вид неустойчивости — типа седла, когда по одному направлению решение устойчиво, а по другому — неустойчиво.

Заметим, что нетрудно предложить химическую реакцию, для которой бифуркационная диаграмма в двумерном пространстве параметров изображается поверхностью с особенностью типа сборки, что приводит к такому явлению, как гистерезис, который может служить моделью химической памяти.

Рассмотренные в этом разделе модели демонстрирует новый подход в химии, состоящий в том, что реакции не всегда предсказуемы (что объясняется существованием неустойчивых состояний в разобранном примере брюсселятора) и не всегда развиваются плавно.

8.6. Самоорганизация в биологических системах

8.6.1. Возникновение жизни. Одним из основных разделов теоретической биологии является теория эволюции, основную идею которой составляет дарвиновский отбор. Однако в ее первоначальном виде идея дарвиновского отбора требует уточнения, в первую очередь касающегося условий отбора и его темпов. так как анализ вероятности спонтанного возникновения жизни на Земле, т.е. вероятности случайного образования простейших живых организмов и их эволюции, крайне мала, по оценкам, она составляет величину порядка 10^{-60} и менее [18]. Эти оценки получаются в предположении равной вероятности группировки атомов или молекул в более сложные соединения в том или ином порядке.

Поэтому основную проблему теории возникновения жизни на Земле можно сформулировать следующим образом: могла ли в эволюции биосферы накопиться информация (например, о строении белковых молекул и т.п.) без маловероятных случайностей?

Рассмотрим возможные пути решения этой проблемы. Согласно современным представлениям, следующим идеям Опарина, Фокса и др., жизнь возникла в первичном бульоне, который до биологической прошел длительную химическую эволюцию. С точки зрения термодинамики, бульон — сильно неравновесная среда, в которой имеются большие градиенты температуры и концентраций, а также богатые свободной энергией химические соединения. В нем уже существовали образовавшиеся химическим путем основные элементы, составляющие современную биосферу. Однако на стадии образования первичного бульона еще не существовало биологической информации, т.е. информации, позволяющей строить организмы, подобные существующему образцу, а также не существовало системы, которая могла бы породить такую информацию, сохранить ее и передать. Вопрос о возникновении такой информации был исследован Эйгеном, который ввел понятие гиперцикла как совокупности каталитических биохимических реакций, в которых репродуцируется белок, катализатором этих реакций служат полинуклеотиды. Гиперциклы являются промежуточной стадией между живыми и неживыми объектами, и среди возможных реакций гиперциклов возможен дарвиновский отбор.

Наиболее тонким моментом теории гиперциклов является вопрос о том, каким именно образом происходит катализ. Дело в том, что вероятность случайного возникновения механизма катализа, работающего на современном этапе биологической эволюции, оценивается невообразимо малой величиной 10^{-7000} [18]. Для сравнения заметим, что все вещество Вселенной эквивалентно 10^{78} атомам водорода, а возраст такого состояния Вселенной, в котором возможны био-

химические реакции, составляет 10^{17} секунд; при этом время на создание или разрушение одной пептидной связи составляет в лучшем случае 10^{-2}–10^{-3} сек. Эти цифры дают наглядное представление о необходимости поиска какого-либо иного механизма этой реакции, осуществляющегося с большей вероятностью. Сейчас намечены лишь общие подходы к созданию моделей таких механизмов.

8.6.2. Математические модели выживания. Итак, на первых этапах эволюции биосферы произошло создание механизм автопродукции и наследования свойств. Однако единого механизма и единого информационного кода для всей биосферы на этом этапе пока еще нет. Он возникает на следующем этапе эволюции, на котором главную роль играет взаимодействие популяций. Математическое моделирование взаимодействия популяций преследует следующие цели:

— исследовать, может ли в результате взаимодействия образоваться "чистая" популяция, обладающая единым вариантом кода, если в начальный момент в системе имеется смешанная популяция, в которой возможны различные гиперциклы с различными вариантами кода;

— изучить, происходит ли в этом процессе выбор одного варианта из нескольких возможных, или же происходит отбор выделенного варианта;

— исследовать помехоустойчивость процесса, т.е. узнать, сколь важную роль играют мутации и возможные ошибки трансляции и репродукции.

Математическая модель, учитывающая конкуренцию за питательную среду между всеми гиперциклами, может быть записана в следующем виде:

$$\begin{aligned} \frac{dx_i}{dt} &= F_i(x_i) - \frac{1}{C_0}\sum_{j=1}^{n} F_j(x_j)x_i, \\ C_0 &= \sum_{j=1}^{n} x_i, \quad i = 1,\ldots,n, \end{aligned} \tag{8.58}$$

где x_i — концентрация гиперциклов i-го вида, $F_i(x_i)$ — скорость производства гиперциклов. Второй член в правой части (8.58) описывает конкурентную борьбу всех особей популяции, в том числе и внутривидовую, описывающуюся слагаемым $x_iF_i(x_i)$. В модели (8.58) общая концентрация принята постоянной, что уменьшает число независимых переменных в (8.58) с n до $n-1$.

В простейшем случае принимают линейную зависимость скорости размножения: $F_i(x_i) = k_ix_i$, где $k_i = \text{const}$, $i = 1,\ldots,n$. Тогда,

обозначив $x_i' = x_i/C_0$, имеем

$$\begin{aligned} \frac{dx_i'}{dt} &= k_i x_i - x_i' \sum_{j=1}^{n} k_j x_j', \\ \sum_{j=1}^{n} x_i' &= 1, \quad i = 1, \ldots, n. \end{aligned} \tag{8.59}$$

Анализ системы (8.59) состоит в поиске особых стационарных точек и исследования их устойчивости методами, рассмотренными ранее в главе 7. Этот анализ показывает, что системы типа (8.59) описывают механизм отбора чистых состояний, т.е. таких, в котором стационарное состояние имеет вид $(1, 0, 0, \ldots, 0)$, т.е. все виды вымирают, кроме первого. К такому результату приходим как в случае большого селективного превосходства первого вида, обеспечивающегося, как правило, большой скоростью роста k_1 (при этом выживание первого вида предопределено и не зависит от начальных условий), так и в случае малого селективного превосходства или даже при отсутствии таковых. В последнем случае выбор выжившего вида не предопределен, а зависит от начального состояния.

8.6.3. Модели роста и взаимодействия популяций.

Модель Мальтуса. Еще одной проблемой биологии, решаемой с помощью математического моделирования, является анализ роста клеточных популяций, популяций особей некоторого вида, народонаселения и др. Основоположником математического популяционного моделирования можно считать Мальтуса, который в конце XVIII века сформулировал законы роста народонаселения в геометрической прогрессии. Этот закон описывается простейшим эволюционным уравнением вида

$$\dot{N} = \alpha N.$$

Очевидно, что такая модель слишком проста, так как не учитывает ограничивающих факторов (ограничения жизненного пространства, ресурсов и т.п.), и может иметь место лишь на достаточно малых промежутках времени. Более полной является модель Ферхюльста, рассмотренная нами ранее и приводящая к логистическому отображению:

$$\dot{N} = \varepsilon(K - N)N.$$

Эта модель предполагает существование некоторого предела K, к которому стремится численность популяции N при $t \to \infty$. Здесь $\mu = \varepsilon K$ — удельная скорость роста популяции, а скорость смертности в популяции пропорциональна ее численности и равна $-\varepsilon N$. Модель Ферхюльста, таким образом, описывает внутривидовую конкуренцию, вызываемую, например, недостатком пищи.

Модель Лотки–Вольтерра. Следующий шаг в математической экологии был сделан в 20-х годах XX века, когда была создана модель взаимодействия двух и более видов. Эти работы были начаты А.Лотки и В.Вольтерра, и получили название моделей типа "хищник–жертва".

Модель Лотки–Вольтерра может быть записана в виде

$$\frac{\partial N_1}{\partial t} = \varepsilon_1 N_1 \left(1 - \frac{\gamma_{12}}{\varepsilon_1} N_2 - \frac{\gamma_{11}}{\varepsilon_1} N_1\right),$$

$$\frac{\partial N_2}{\partial t} = \varepsilon_2 N_2 \left(1 - \frac{\gamma_{21}}{\varepsilon_2} N_1 - \frac{\gamma_{22}}{\varepsilon_2} N_2\right),$$

здесь первые члены в правых частях описывают размножение видов, вторые — конкурентную борьбу, а третьи — самоограничение. Уравнение

$$\frac{\partial N_1}{\partial t} = \frac{\partial N_2}{\partial t} = 0, \tag{8.60}$$

определяющее стационарные точки системы Лотки–Вольтерра, приводит к необходимости вымирания одного из видов, так как стационарной устойчивой точкой при невырожденных условиях (8.60) является точка $(N_1^0, 0)$, либо точка $(0, N_2^0)$. Вырождение (8.60) означает, что популяции находятся в равных условиях, лишь в этом случае возможно существование двух видов.

Выживание двух видов возможно в системе, описываемой уравнениями

$$\frac{\partial N_1}{\partial t} = \alpha_1 N_1 - \alpha N_1 N_2,$$

$$\frac{\partial N_2}{\partial t} = \beta N_1 N_2 - 2\kappa_2 N_2.$$

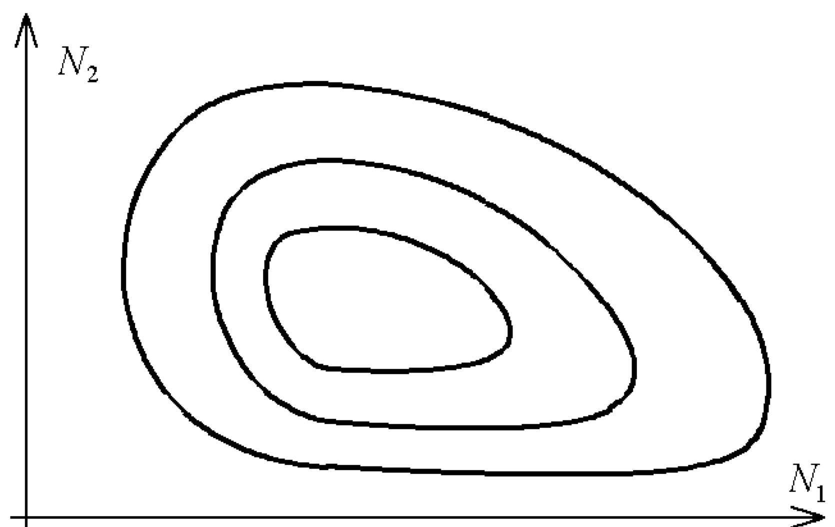


Рис. 8.11. Фазовый портрет динамической системы Лотки - Вольтерра

Эта модель хорошо описывает количество рыб в Адриатическом море, число рысей и зайцев и др. Она демонстрирует циклическое поведение численности популяций N_1 и N_2 во времени, фазовый портрет системы приведен на рис. 8.11.

Модели с обострением. В экологических моделях часто возникают ситуации, когда численность особей популяции растет не по линейному закону, а быстрее. Например, размножение определяется парными столкновениями, моделирующими встречу женской и мужской особей, численности же как тех, так и других пропорциональны общему числу особей. В этом случае модель задается уравнением вида

$$\dot{n} = \alpha n^2. \tag{8.61}$$

Общее решение уравнения (8.61) имеет вид

$$n(t) = \frac{1}{\alpha(t-c)}, \tag{8.62}$$

и изображено на рис. 8.12; для него характерно, что n обращается в бесконечность за конечное время. Решения вида (8.62) описывают так называемые режимы с обострением. В частности, моделью (8.61) с хорошей точностью описывается рост народонаселения Земли со значением константы c в (8.62), равным 2030, т.е. эта модель к 2030-му году предсказывает бесконечно большое число людей на Земле. Естественно, что реально бесконечность не является достижимой, в окрестности точки $t = c$ начинают проявляться процессы, ограничивающие рост населения. Например, это может описываться зависимостью α от n. Однако, сами режимы с обострением обусловливают существование двух временных областей — одна характеризуется бесконечным ростом решения, а другая — его затуханием, причем переброска из одного состояния в другое может быть осуществлена за счет флуктуаций.

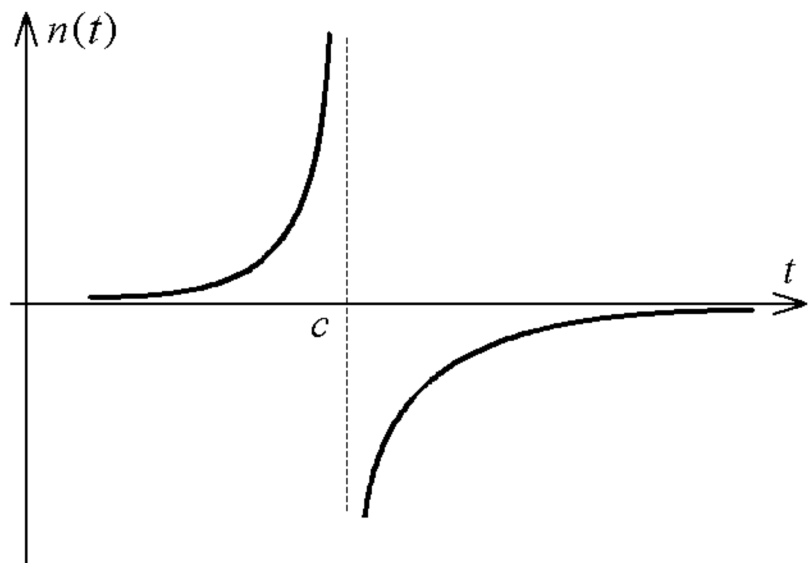


Рис. 8.12. Режим с обострением

8.6.4. О модели морфогенеза. Рассмотрим еще одну математическую модель, описывающую биологический процесс морфогенеза. Этот процесс может быть пояснен на примере гидры. Если удалить часть головной области гидры и пересадить удаленные клетки на другое место ее тела, то в зависимости от удаленности места пересадки от головы из пересаженных клеток развивается или не развивается новая голова. Если место пересадки достаточно близко к голове, то рост новой головы подавляется, в противном случае вырастает новая голова. Это наводит на мысль, что существует по крайней мере два вещества — активатор и ингибитор, одно из них (активатор) стимулирует рост новой головы, а другое (ингибитор) — подавляет; результат опыта может быть обусловлен пространственным распределением этих веществ в теле гидры.

Пусть $a(x,t)$ и $h(x,t)$ — концентрации активатора и ингибитора, соответственно. Одномерная математическая модель морфогенеза записывается в виде системы дифференциальных уравнений в частных

производных:

$$\frac{\partial a}{\partial t} = \rho + k\frac{a^2}{h} - \mu a + D_a\frac{\partial^2 a}{\partial x^2},$$
$$\frac{\partial h}{\partial t} = ca^2 - \nu h + D_h\frac{\partial^2 h}{\partial x^2}, \tag{8.63}$$

здесь ρ описывает генерацию, $-\mu a$ — распад активатора, подавление скорости роста концентрации активатора при больших содержаниях ингибитора описывается членом ka^2/h, аналогично рост концентрации ингибитора за счет активатора и его распад описываются членами ca^2 и νh в (8.63), слагаемые $D_a(\partial^2 a/\partial x^2)$ и $D_h(\partial^2 h/\partial x^2)$ описывают распространение активатора и ингибитора за счет диффузии.

Стандартным образом определяются стационарные однородные по пространству состояния системы (8.63), т.е. как не зависящие от x решения уравнения $\partial a/\partial t = \partial h/\partial t = 0$, и исследуется их устойчивость путем анализа собственных значений оператора линеаризации правой части (8.63). При этом следует иметь в виду, что оператор линеаризации содержит оператор Лапласа. Решение задачи на собственные значения линеаризованного оператора правой части (8.63) позволяет найти условие появления неустойчивости (т.е. роста) пространственной структуры, описываемой собственным вектором линеаризованного оператора с собственным числом, реальная часть которого неотрицательна, а мнимая — равна нулю. Характерное поведение структур, появляющихся как решение уравнения (8.63), есть пространственное чередование пиков концентраций активатора, появление новой структуры вблизи пика подавляется высокой концентрацией ингибитора.

8.7. Клеточные автоматы

8.7.1. Математическая модель клеточного автомата. Знакомство с характерным поведением активных сред и диссипативных структур позволяет целенаправленно создавать нелинейные динамические системы, обладающие свойствами, моделирующими реальные и достаточно сложные объекты. В частности, из простых элементов, меняющих свое состояние в дискретные моменты времени по определенному закону в зависимости от того, в каком состоянии находился сам элемент и его ближайшее окружение в предыдущий момент времени, можно собрать среду, отражающую те или иные особенности физических объектов, что позволяет моделировать поведение упругих сред, явления гидродинамики, кинетики, моделировать деятельность человеческого мозга по переработке информации, заключающейся, в

частности, в узнавании образов, извлечении ассоциаций и др. Сети, сделанные из таких элементов, называются клеточными автоматами.

Подход к моделированию физической реальности с помощью описания шагов клеточного автомата составляет так называемый алгоритмический подход; он значительно отличается от привычного, сводящегося, как правило, к дифференциальным уравнениям; однако при численном решении дифференциальных уравнений эти подходы смыкаются. Если же удастся сформулировать адекватное описание природы в терминах алгоритмов, то это существенно изменит нашу картину мира и упростит ее моделирование и изучение с помощью вычислительных машин.

Обратимся к математической модели клеточного автомата. Пусть задана система, состоящая из N элементов. Пронумеруем их от 1 до N и зададим между ними связи, указав для каждого элемента, какие из элементов системы являются его соседями, иными словами — зададим топологию системы. Окрестность j-го элемента, состоящую из его соседей, обозначим $O(j)$. Состояние j-го элемента в n-й момент времени t_n обозначим $a_j^{(n)}$ и будем считать, что $a_j^{(n)}$ может быть как числом, так и вектором и принимать одно значение из некоторого заданного набора. Правило перехода от состояния $a_j^{(n)}$ в момент времени t_n в состояние $a_j^{(n+1)}$ в следующий момент времени t_{n+1} зададим формулой

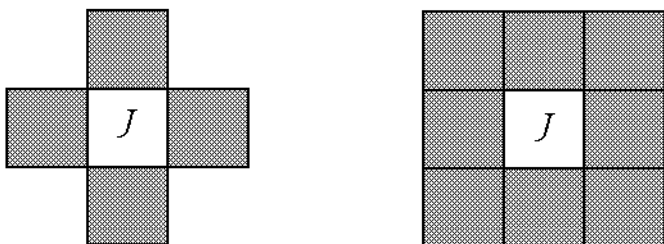


Рис. 8.13. Различные расположения соседних элементов на плоскости

$$a_j^{(n+1)} = F\left(a_j^{(n)}, \sum_{i\in O(j)} a_i^{(n)}\right), \qquad (8.64)$$

т.е. состояние j-го элемента в момент времени t_{n+1} определяется его состоянием в момент времени t_n и суммой состояний его соседей в этот же момент времени. При таком определении клеточный автомат не обладает памятью, и для того, чтобы сконструировать клеточный автомат с памятью, нужно, чтобы его состояние определялось и значениями $a_j^{(n-1)}, a_j^{(n-2)}, \ldots, a_i^{(n-1)}, a_i^{(n-2)}, \ldots, i \in O(j)$. Клеточный автомат, описываемый соотношением (8.64), называется детерминированным. Для того чтобы задать случайный клеточный автомат, надо определить набор переходных вероятностей

$$w\left(a_j^{(n+1)} \middle| a_j^{(n)}, \sum_{i\in O(j)} a_i^{(n)}\right),$$

задающих вероятность найти j-й элемент в $(n+1)$-й момент времени t_{n+1} в состоянии $a_j^{(n+1)}$ при условии, что в предыдущий момент времени он находился в состоянии $a_j^{(n)}$, а сумма состояний его соседей равнялась $\sum\limits_{i \in O(j)} a_i^{(n)}$.

Заметим, что окрестность элемента может быть определена разными способами, некоторые из них показаны на рис. 8.13. Кроме того, подчеркнем, что набор соседей не обязательно связан с топологией пространства, в которое погружена сеть элементов клеточного автомата. Например, на плоскости можно разместить сеть с топологией трехмерного пространства; для этого вообразим "трехмерную рыболовную сеть", сплетенную из гибких ниток, и постелем эту сеть на двумерную поверхность так, как это изображено на рис. 8.14.

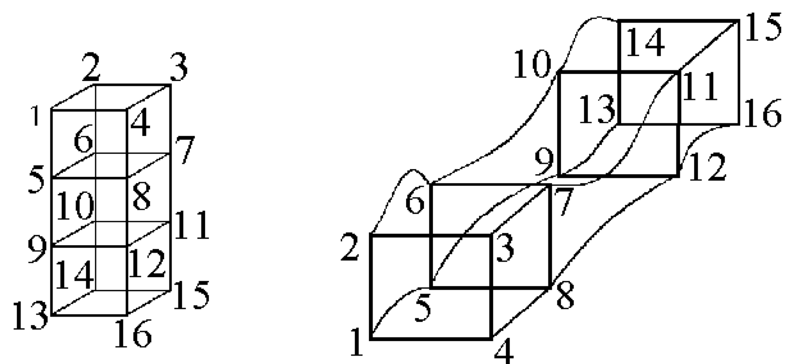


Рис. 8.14. Топология клеточного автомата в пространстве и на плоскости

Подсчитаем общее число возможных моделей детерминированных клеточных автоматов с фиксированным числом соседей и заданным числом состояний. Пусть чисто возможных состояний элемента равно K и задается числами $0, 1, \ldots, K-1$, а сам элемент может иметь r соседей. Согласно (8.64), значение состояния $a_j^{(n+1)}$ определяется комбинацией двух чисел — состоянием $a_j^{(n)}$ элемента в момент времени t_n и суммой $\sum\limits_{i \in O(j)} a_i^{(n)}$.

Первое из них может принимать K значений, а второе — Kr значений, итого rK^2 возможных комбинаций. Каждой паре этих чисел можно сопоставить любое из K значений $a_j^{(n+1)}$, значит, количество различных клеточных автоматов при фиксированных r и K равно K^{rK^2}, что при $r = 4$, $K = 2$ составляет $K^{rK^2} = 256$, а при $r = 8$, $K = 2$ — соответственно $K^{rK^2} = 2^{32} = 65\,636$.

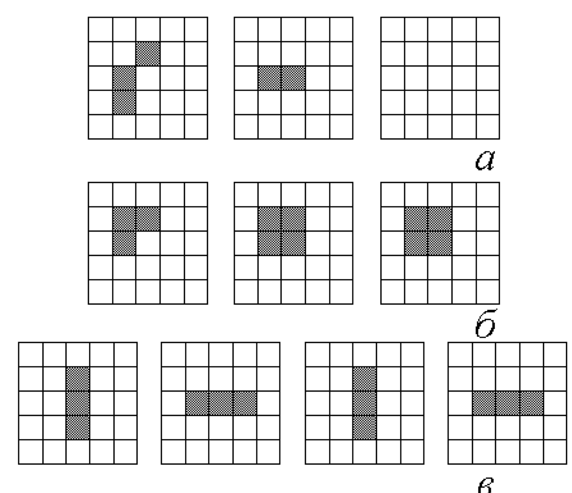


Рис. 8.15. Эволюция структур в игре "Жизнь": исчезновение структуры (*а*); стабилизация (*б*); "колебания" (*в*)

Динамика клеточного автомата необратима, так как каждой из rK^2 различных комбинаций аргументов функции $F(\cdot)$ в (8.64) ставится в соответствие лишь K возможных состояний $a_j^{(n+1)}$, и множество состояний, породивших

некоторое значение $a_j^{(n+1)}$, не может всегда состоять из единственного набора $a_j^{(n)}$, $a_i^{(n)}$, $i \in O(j)$.

Клеточные автоматы можно разбить на четыре класса. К первому относятся те клеточные автоматы, которые за конечное число шагов достигают определенного состояния. Примеры таких автоматов рассмотрены в разделе 8.3, посвященной моделированию активных сред, составленных из бистабильных или из возбудимых элементов. Эти состояния достигаются не зависимо от начального состояния. Примером такого автомата может служить одномерная цепочка ячеек с двумя соседями, число возможных состояний каждой ячейки равно двум, и переход ячейки из одного состояния в другой происходит согласно формуле

$$a_j^{(n+1)} = \begin{cases} 1, & \text{если } a_j^{(n)} = a_{j-1}^{(n)} = a_{j+1}^{(n)} = 1, \\ 0 & \text{в противном случае.} \end{cases}$$

Легко видеть, что все элементы клеточного автомата в конце концов перейдут в состояние $a = 0$.

Клеточные автоматы второго класса генерируют последовательность простых структур, которые могут быть стационарными или периодическими во времени. Например, если в предыдущем примере правило перехода задается формулой

$$a_j^{(n+1)} = \begin{cases} 1, & \text{если } a_j^{(n)} = 1,\ a_{j-1}^{(n)} + a_{j+1}^{(n)} = 0 \text{ или } 1, \\ \\ 1, & \text{если } a_j^{(n)} = 0 \text{ или } a_{j-1}^{(n)} + a_{j+1}^{(n)} = 2, \\ 0 & \text{в противном случае,} \end{cases}$$

т.е. элемент сохраняет состояние активности ($a = 1$), если среди его соседей справа и слева не более одного активного элемента, и становится активным, если оба его соседа активны.

Клеточные автоматы III и IV классов демонстрируют сложное динамическое поведение: при $n \to \infty$ различные карты активности сменяют друг друга, ни разу не повторяясь. Локальные возмущения порождают процесс изменения активности, захватывающий со временем все большую часть сети. Автоматы III класса обладают эргодичностью, т.е. по прошествии достаточно большого промежутка времени с начала эволюции сеть забывает о своих начальных значениях, динамика же клеточных автоматов IV класса существенно зависит от начальных условий.

8.7.2. Игра "Жизнь". Примером клеточного автомата IV класса является известная игра "Жизнь", предложенная Джоном Конуэем в качестве математического развлечения. В клеточном автомате

"Жизнь" правила таковы: каждый элемент находится в состоянии покоя или активности, он переходит из состояния покоя в состояние активности, если по соседству с ним оказалось ровно три активных элемента, он сохраняет состояние активности, если среди соседей есть 2 или 3 активных элемента. Число соседей при этом равно восьми.

Игра "Жизнь" демонстрирует разнообразное поведение в зависимости от начального состояния. Например, некоторые структуры исчезают, как показано на рис. 8.15,*а*, а некоторые достигают стационарного поведения, рис. 8.15,*б*, *в*. Есть конструкции, которые движутся и тем самым напоминают бегущий импульс в активной среде, к ним относится так называемый планер, или парусник, изображенный на рис. 8.16. Он превращается в первоначальную фигуру через четыре этапа и при этом смещается на один элемент вниз и на один элемент вправо. На рис. 8.16 показаны последовательные состояния планера и направления его смещения. Есть и более сложные конфигурации, например, планерное ружье, представляющее собой пространственный осциллятор, который через 30 поколений элементов возвращается в исходное состояние и при этом испускает один планер; есть "пожиратель планеров" — конструкция, которая поглощает планеры и не изменяется.

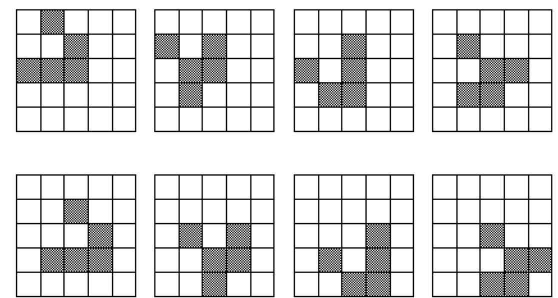


Рис. 8.16. Эволюция "планера"

8.7.3. Фильтр движущихся целей. Одно из интересных свойств клеточных автоматов состоит в том, что с их помощью можно создать системы, отфильтровывающие изображения движущихся объектов от шумов. Например, пусть клеточный автомат состоит из одномерной цепочки (линейки) элементов, и на каждый j-й элемент линейки в моменты времени $t_1, t_2, \ldots$ поступает сигнал (изображение) I_j^n, $j = 0, 1, \ldots$, $n = 1, 2, \ldots$ При фиксированном n изображение I_j^n представляет собой последовательность нулей и единиц, которые являются зашумленными изображениями нескольких смещающихся объектов, причем за единицу времени изображение объекта может смещаться на один элемент влево или вправо, и проблема состоит в том, чтобы подавить шум и выделить движущиеся изображения.

Для решения этой задачи состояние каждого элемента будем описывать тремя целыми числами — R_j^n, L_j^n и S_j^n, причем S_j^n принимает значение либо 0, либо 1, а R_j^n и L_j^n — любое значение из множества $0, 1, \ldots, M$. Правила перехода даются следующими соотношениями:

$$L_j^{(n+1)} = \begin{cases} L_{j+1}^{(n)} + (1 - S_j^{(n)}), & \text{если } I_j^{(n+1)} = 1, \\ L_{j+1}^{(n)} - 1, & \text{если } I_j^{(n+1)} = 0, \end{cases}$$

$$R_j^{(n+1)} = \begin{cases} R_{j-1}^{(n)} + (1 - S_j^{(n)}), & \text{если } I_j^{(n+1)} = 1, \\ R_{j-1}^{(n)} - 1, & \text{если } I_j^{(n+1)} = 0, \end{cases}$$

$$S_j^{(n+1)} = S_j^{(n+1)}.$$

и если вычисленные по этим правилам значения $R_j^{(n+1)}$ или $L_j^{(n+1)}$ оказываются больше M, то их полагают равными M, аналогично $R_j^{(n+1)} = 0$ или $L_j^{(n+1)} = 0$, если $R_j^{(n+1)}$ или $L_j^{(n+1)}$ оказываются в результате вычислений отрицательными. Таким образом, в цепочке слева направо передаются сигналы $R_j^{(n)}$, а справа налево — сигналы $L_j^{(n)}$. Передавая сигналы, элемент:

— увеличивает его значение на единицу, если в начальный момент этот элемент освещен (т.е. $I_j^{(n+1)} = 1$), а в предыдущий момент был темным ($I_j^{(n)} = 1$);

— сохраняет значение сигнала неизменным, если он был освещен в предыдущий момент и остался освещенным в настоящем;

— уменьшает сигнал на единицу, если он не оказался освещенным в настоящий момент времени.

Если процесс продолжается достаточно долго, то по состоянию автомата можно определить координаты и направления движения объектов, так как при усреднении шумы усреднятся, а сигналы накопятся.

С помощью клеточных автоматов можно моделировать процессы гидродинамики, колебаний упругих вязких сред и другие, следует, однако, иметь в виду, что в клеточных автоматах из-за наличия дискретной решетки элементов не выполняются некоторые физические законы, такие, например, как инвариантность всех инерциальных систем отсчета (галилеевская инвариантность), изотропность пространства и т.п.

8.7.4. Клеточная модель физической реальности. В работах С.Я. Берковича клеточный автомат, составленный из автоколебательных элементов (осцилляторов), используется для алгоритмического описания физической реальности. В рассматриваемой им модели фаза колебания $\vartheta_j(t_{n+1})$ элемента с номером j в момент времени t_{n+1} связан с фазами колебания элементов из его окрестности соотношением

$$\vartheta_j(t_{n+1}) = \sum_{i \in O(j)} a_{ji} \vartheta_j(t_n), \tag{8.65}$$

здесь окрестность $O(j)$ содержит и сам элемент j; коэффициент "самовлияния" a_{jj} полагается близким к единице: $0 < 1 - a_{jj} \ll 1$, и при удалении от элемента с номером j коэффициенты влияния быстро убывают.

Уравнение (8.65) может быть аппроксимировано дифференциальным уравнением. Принимая во внимание лишь ближайших соседей и учитывая симметричный характер взаимодействия элементов, запишем простейшее уравнение для фаз колебаний в виде

$$\frac{\partial\vartheta}{\partial t} = -b\vartheta + D_0\Delta\vartheta,$$

где ϑ — оператор Лапласа, D_0 — коэффициент диффузии. При подстановке $\vartheta = \overline{\vartheta}e^{-bt}$ это уравнение приводится к стандартному уравнению теплопроводности

$$\frac{\partial\overline{\vartheta}}{\partial t} = D_0\Delta\overline{\vartheta},$$

решение которого для точечного мгновенного источника известно. Итак,

$$\vartheta = \frac{1}{(2D_0\sqrt{\pi t})^3}e^{\left(-\frac{R^2}{4D_0T}-bt\right)}.$$

Более общее уравнение для фазы, учитывающее "дальних" соседей, можно записать в виде

$$\frac{\partial\vartheta}{\partial t} = -b\vartheta + D_0\Delta\vartheta + D_1\Delta^2\vartheta,$$

где Δ^2 — дифференциальный оператор четвертого порядка. Найдем его решение в виде волны, распространяющейся вдоль оси z в цилиндрической системе координат, т.е. в виде функции $\vartheta = f(z-vt)$, откуда для f получим уравнение

$$-vf'(s) = -bf(s) + D_0f''(s) + D_1f''''(s), \quad s = z - vt. \tag{8.66}$$

Будем искать решение (8.66) в виде $f = e^{\lambda t}$, тогда для λ получим характеристическое уравнение

$$D_1\lambda^4 + D_0\lambda^2 + v\lambda - b = 0. \tag{8.67}$$

Если $D_1 \ll D_0$, то два корня (8.67) могут быть оценены из приближенного квадратного уравнения

$$D_0\lambda^2 + v\lambda - b = 0;$$

с другой стороны, оставляя члены только второй и четвертой степени λ, получим второе уравнение для оценки корней (8.67) в виде

$$D_1\lambda^4 + D_0\lambda^2 = 0.$$

Итак, оценки корней (8.67) имеют вид

$$\lambda_1 = \frac{b}{v}, \quad \lambda_2 = -\frac{v}{D_0}, \quad \lambda_{34} = \pm i\sqrt{\frac{D_0}{D_1}},$$

и общее решение (8.66) может быть оценено формулой

$$f(s) = A_1 e^{\lambda_1 s} + A_2 e^{\lambda_2 s} + A_3 \cos(|\lambda_{34}| s + \gamma_0), \tag{8.68}$$

где константы A_1, A_2, A_3, γ_0 определяются из начальных или граничных условий. В рассматриваемом случае эти условия должны обеспечивать существование волны, в цилиндрической системе координат совершающей одновременно поступательное и вращательное движение. Радиус r и угловая скорость вращения фазы $\dot{\varphi}$ определяются С.Я. Берковичем из условия, что за конечное число шагов фаза совершает целое число оборотов.

Распространение стационарного решения (8.68) вызывает возмущение фаз в окрестных элементах и порождает структуры, называемые Берковичем синхроформации; они интерпретируются им как элементарные частицы — электроны, протоны, нейтроны, нейтрино и т.п., при этом дискретный набор частиц связан с условием соизмеримости периодов волн (8.68). Разные типы слагаемых в (8.68) связываются с различными частицами: протон — с возрастающей экспонентой, электрон — с убывающей; их линейная комбинация дает нейтрон, а мнимые экспоненты — нейтрино и т.п.

Концепция Берковича интересна тем, что позволяет взглянуть на описание физической реальности с новой точки зрения, без привычных стереотипов, и дает пример попытки принципиально нового подхода к пониманию законов природы.

8.8. Обучающиеся системы

8.8.1. Модель Изинга. Сеть дискретных элементов, связанных между собой по определенным законам, может служить моделью искусственного интеллекта в том смысле, что такая динамическая система может проявлять свойства, характерные для высокоразвитого интеллекта, такие, например, как ассоциативная память, узнавание сложных образов и т.п. В частности, конструкция динамической системы, демонстрирующей ассоциативную память, может быть основана на следующей идее. Пусть в системе имеется набор аттракторов, с каждым из которых связана типичная картина–образ, и все пространство состояний (фазовое пространство) делится на области притяжения так, что любое начальное состояние (кроме состояний

на сепаратрисах, которые, впрочем, неустойчивы) эволюционируют к какому-либо аттрактору. Тогда с любым предъявленным образом ассоциативно связан один из заданных образов, а именно тот, к области притяжения которого принадлежит предъявленный образ.

Рассмотрим реализацию такой системы, предложенную Дж. Хопфилдом на базе специальных физических систем, носящих название спиновых стекол. В этой системе имеется N элементов (так называемых изинговских спинов), каждый из которых может находиться в двух состояниях — $S_i = +1$ и $S_i = -1$. Спины взаимодействуют между собой так, что энергия парных взаимодействий i-го и j-го спинов равна $(-J_{ij}S_iS_j)$, а полная энергия системы складывается из энергий парных взаимодействий и равна

$$E = -\frac{1}{2}\sum_{i,j=1}^{N} J_{ij}S_iS_j = -\frac{1}{2}(\widehat{J}\mathbf{S}, \mathbf{S}), \tag{8.69}$$

где $\widehat{J}$ — матрица взаимодействий. Эволюция системы сводится к тому, что полная энергия взаимодействий E стремится к минимуму.

Если все $J_{ij} > 0$, то в состоянии минимума энергии все спины направлены в одну сторону, т.е. либо $S_i = 1$, $i = +1, \ldots, N$, либо $S_i = -1$, $i = 1, \ldots, N$, так как при этом энергии всех парных взаимодействий достигают своих минимальных значений, равных $-J_{ij}S_iS_j$. Такое состояние называется ферромагнитным.

В спиновых стеклах матрица $\widehat{J}$ имеет как положительные, так и отрицательные элементы, что приводит к тому, что в состоянии минимума полной энергии все парные взаимодействия не могут иметь одновременно минимальную энергию, так как обязательно найдутся три элемента, для которых энергия их взаимодействия $J_{ij}S_iS_j + J_{jk}S_jS_k + {}$ ${}+ J_{ki}S_kS_i$ в состоянии минимума имеет отрицательные слагаемые. В такой системе имеется несколько состояний с локальным минимумом полной энергии — это устойчивые стационарные состояния.

Рассмотрим, как можно “записать” образ в такой системе. Пусть матрица $\widehat{J}$ симметрична, и запоминаемый образ, который система должна воспроизводить при эволюции, имеет распределение спинов, описываемое вектором $\mathbf{S}_1$. Зададим матрицу $\widehat{J}_1$ ранга единица по формуле

$$J_{1ij} = \frac{1}{N}S_{1i}S_{1j}. \tag{10}$$

Это — матрица проецирования на вектор $\mathbf{S}_1$. Согласно (8.69) система будет иметь минимальную энергию тогда, когда $\mathbf{S}$ есть собственный вектор матрицы $\widehat{J}$, отвечающий максимальному собственному значению (если он имеет координаты, равные или единице, или

минус единице). Но единственный собственный вектор матрицы (10), отвечающий ненулевому собственному значению, и есть $\mathbf{S}_1$, значит, создав спиновое стекло с матрицей взаимодействия (10), мы зададим систему, "помнящую" образ $\mathbf{S}_1$, т.е. эволюционирующую к нему, если задать начальное распределение спинов из некоторой окрестности $\mathbf{S}_1$.

Алгоритм переходов системы из одного состояния в другое можно задать следующим образом: рассмотрим j-й элемент, и вычислим "поле" $h_j = \sum_{i=1}^{N} J_{ij} S_i$, создаваемое всеми другими элементами. Энергия взаимодействия этого элемента с "полем" равна произведению величины "поля" h_j на спин S_j, взятому с обратным знаком. Поэтому, если сомножители имеют один знак (спин направлен вдоль "поля"), то эта энергия минимальна, в противном случае она может быть уменьшена поворотом спина. Итак, эволюция системы состоит в последовательном развороте всех спинов вдоль поля, создаваемого соседними элементами.

Таблица 8.1

1	John Steward Denker 8128
2	Lawrence David Jacke 7773
3	Richard Edwin Howard 5952
4	Wayne P. Hubbard 7077
5	Brian W. Straughn 3126
6	John Henry Scofield 8109

На практике часто требуется записать несколько образов. Эта задача решается аналогично предыдущей, если все образы, подлежащие запоминанию, ортогональны. Действительно, если взять матрицу проецирования на векторы $\mathbf{S}_1, \ldots, \mathbf{S}_M$, т.е. построенную согласно формуле

$$J_{Mij} = \frac{1}{N} \sum_{k=1}^{M} S_{ki} S_{kj}. \tag{11}$$

то любое состояние $\mathbf{S} = \mathbf{S}_k$, $k = 1, \ldots, M$, является состоянием локального минимума полной энергии, так как при этом все спины направлены по полю. Действительно, матрица $\widehat{J}_M$ не меняет направление вектора $\mathbf{S}_k$, так как он по построению является ее собственным вектором, и разворот любого спина в другом направлении приведет к росту энергии системы.

В реальных системах, помимо запомненных согласно (11) образов, существуют и другие состояния с локальным минимумом энергии — так называемые фантомы. Такими фантомами, в частности, являются образы $\mathbf{S} = -\mathbf{S}_k$, $k = 1, \ldots, M$, т.е. противоположно направленные по отношению к заданным. Возможны и иные состояния.

Если исходные образы не ортогональны, построение матрицы взаимодействий по формуле (11) приведет к искажениям запомненных образов. Однако если размерность N векторов $\mathbf{S}$ велика, то два случайно выбранные векторы $\mathbf{S}_1$ и $\mathbf{S}_2$ "почти ортогональны", так как

в силу закона больших чисел

$$(\mathbf{S}_1, \mathbf{S}_2) = \frac{1}{N} \sum_{j=1}^{N} S_{1j} S_{2j} \to 0$$

при $N \to \infty$, т.е. при больших N и малых M эти искажения могут оказаться незначительными.

Проиллюстрируем принцип действия динамической системы с ассоциативной памятью, построенной согласно модели Хопфилда, на примере, в котором в качестве "образов" были взяты двоичные последовательности, кодирующие фамилии и имена телефонов шести сотрудников фирмы Bell, приведенные в табл. 8.1 [8]. В табл. 8.2 показано, как выглядит временная эволюция системы при подаче на вход последовательности символов "John S" и "John H". Там же приведен пример фантома, к которому сходится эволюция системы при подаче на вход "мусора".

Таблица 8.2

0	0	john s
20	-0.0784	john sdewirtbnenimw 8109
40	-0.8426	john sdewirtbnenimw 8129
60	-0.8451	john sdewirtbnenimv 8129
80	-0.8581	john sdewirt nenkmv 8128
100	-0.9099	john sdewart denker 8128
120	-0.9824	john steward denker 8128
0	0	john h
20	-0.0665	john hdnwybqbofmalt 8109
40	-0.8579	john henwydqcofield 8109
80	-0.9094	john henry scofield 8109
0	0	garbage
20	-0.0244	garbagee lafj naabd 5173
40	-0.6280	garbaged derjd naabd 7173
80	-0.7595	gasbabed derjd naabd 7173
140	-0.8264	fasjebad derjd naabd 7173

В табл. 8.2 в первом столбце приведено время (в условных единицах), прошедшее с момента начала эволюции системы, во втором столбце — энергия системы, а в третьем — состояние системы.

Видно, что система достаточно быстро "по ассоциации" восстанавливает полностью фамилию и телефон сотрудника, однако существует и "фантом" — бессмысленный набор знаков, к которому эволюционирует система от начального состояния "garbage". Интересно, что даже по искаженному имени "Larry", представленному как начальный образ, система правильно восстанавливает полное имя и телефон сотрудника "Lawrence David Jacke", числящегося в табл. 8.1 под номером 2 (см. табл. 8.3).

Таблица 8.3

0	0	larry
20	-0.0213	larrynce david jacke 7773
80	-0.8813	lasrence david jacke 7773
120	-1.001	lawrence david jacke 7773

Рассмотренная здесь модель Хопфилда является простейшим примером обучающейся системы. Недостатком ее является то, что она не дает возможности управлять процессом распознавания, т.е. изменять области притяжения аттракторов, ассоциирующихся с запомненными образами. Более сложные распознающие системы выделяют существенные признаки образа, тем самым решают задачу поиска закономерностей во входных данных, свойственную высокоразвитому интеллекту.

8.8.2. Нейронные сети. Одним из первых распознающий устройств является перцептрон, появившийся в шестидесятых годах. Его простейшая модель состоит из двух слоев бистабильных элементов, рис. 8.17.

На входной слой подается двоичный сигнал I_j, $j = 1, \ldots, N$, а соответствующий ему выходной сигнал характеризуется двоичной последовательностью O_k, $k = 1, \ldots, n$. Каждый входной элемент с номером j соединен с несколькими (возможно, и со всеми) выходными связью с весовым коэффициентом J_{jk}, и состояние выходного элемента с номером k определяется по правилу

$$O_k = \vartheta(a_k - a_{0,k}), \quad \text{где} \quad a_k = \sum_{j=1}^{N} J_{kj} I_j,$$

а $\vartheta(\cdot)$ — ступенчатая функция:

$$\vartheta(x) = \begin{cases} 1, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Таким образом, картина активности выходных элементов связана с картиной (изображением), подаваемым на входные элементы, и такие системы реагируют определенным образом на входные воздействия. По принципу перцептрона построена, например, система зрительного восприятия лягушки — благодаря специальным связям между рецепторами и нервными клетками лягушка реагирует на движущиеся объекты.

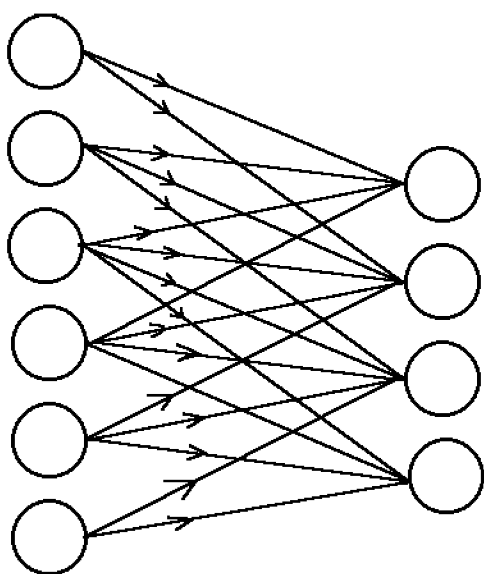


Рис. 8.17. Структура нейронной сети

Выработка реакции системы на предъявляемые ей образы осуществляется в процессе обучения и состоит в выборе таких коэффициентов связи J_{kj}, чтобы при подаче на вход некоторого множества картин $\{I_j\}$ на выходе формировались желаемые картины активности — т.е. заданные образы, разбивающие, в каком-то смысле, множество входных картин на подмножества, каждое подмножество при

этом связывается с одним образом, отвечающим выходной картине активности. Сравнивая выходную картину $\{O_k\}$ с желаемой $\{D_k\}$ при фиксированном сигнале на входе, определим сигнал ошибки по формуле $E_k = D_k - O_k$, $k = 1, \ldots, n$. Зная сигнал ошибки, можно подобрать значения коэффициентов связи J_{kj} так, чтобы уменьшить эту ошибку, например, согласно алгоритму

$$\Delta J_{kj} = \epsilon E_k I_j, \tag{8.70}$$

где ϵ — малый параметр, управляющий скоростью обучения. Действительно, так как E_k и I_j могут принимать значения либо 0, либо 1, то при изменении коэффициентов связи J_{kj} согласно (8.70) можно только уменьшить модуль сигнала ошибки, так как $\Delta J_{kj} \neq 0$ тогда и только тогда, когда одновременно состояние O_k ошибочно и j-й входной элемент активен.

Двухслойный перцептрон достаточно прост, и его возможности довольно ограничены, в частности, на нем невозможно воспроизвести функцию "исключительного ИЛИ", которая заключается в том, что только при подаче активного сигнала лишь на один из двух входных элементов A и B, соединенного с выходным элементом C, последний оказывается в активном состоянии. Действительно, правило "исключительного ИЛИ" записывается в виде

$$D_C = \begin{cases} 0, & \text{если} \quad I_A = 0,\ I_B = 0, \\ 1, & \text{если} \quad I_A = 1,\ I_B = 0, \\ 1, & \text{если} \quad I_A = 0,\ I_B = 1, \\ 0, & \text{если} \quad I_A = 1,\ I_B = 1. \end{cases} \tag{8.71}$$

Так как, в соответствии с моделью двухслойного перцептрона, выполнено $O_C = \vartheta(J_A I_A + J_B I_B - a_0)$, то

$$O_C = \begin{cases} \vartheta(-a_0), & \text{если} \quad I_A = 0,\ I_B = 0, \\ \vartheta(J_A I_A - a_0), & \text{если} \quad I_A = 1,\ I_B = 0, \\ \vartheta(J_B I_B - a_0), & \text{если} \quad I_A = 0,\ I_B = 1, \\ \vartheta(J_A I_A + J_B I_B - a_0), & \text{если} \quad I_A = 1,\ I_B = 1. \end{cases}$$

Для того чтобы $\vartheta(-a_0) = 0$, а $\vartheta(J_A I_A - a_0) = \vartheta(J_B I_B - a_0) = 1$, как того требует (8.71), надо, чтобы $0 < a_0 < \min(J_A, J_B)$, но тогда, в силу неубывания функции $\vartheta(\cdot)$, должно выполняться равенство $\vartheta(J_A I_A + J_B I_B - a_0) = 1$, что не согласуется с (8.71).

Для преодоления этого и других ограничений были предложены многослойные перцептроны, в которых между слоем входных и выходных элементов расположены еще один или несколько "скрытых" слоев. Для их обучения широко используется так называемый метод обратного распространения ошибок, который по своей сути является

градиентным методом минимизации ошибки распознавания. Рассмотрим систему, состоящую из N слоев, причем число элементов N_j, $j = 1, \ldots, N$ в каждом слое может быть различным. Первый слой состоит из входных элементов, последний с номером N — из выходных, и между ними расположены $N - 2$ скрытых слоя; связи существуют только между элементами соседних слоев. При поступлении некоторого сигнала $x_i^{(n)}$ на вход i-го элемента n-го слоя этот элемент вырабатывает выходной сигнал $y_i^{(n)}$, равный

$$y_i^{(n)} = \frac{1}{1 + \exp(-x_i^{(n)})}, \tag{8.72}$$

а сигнал на входе j-го элемента следующего слоя формируется как взвешенная сумма выходных элементов предыдущего слоя:

$$x_j^{(n+1)} = \sum_i J_{ji}^{(n)} y_i^{(n)}, \tag{8.73}$$

коэффициенты связи $J_{ji}^{(n)}$ могут быть как положительными, так и отрицательными. Задача обучения состоит в том, чтобы при подаче на входной слой элементов изображения из некоторого класса $\mathcal{L}_j$ сигнал на выходе соответствовал некоторой заданной картине (образу) D_j, отвечающему классу $\mathcal{L}_j$ входных сигналов.

Более формально, пусть задано M различных эталонных изображений $\{I_j, j = 1, \ldots, N_{\text{вх}}\}_m$, для каждого из которых известна характерная реакция $\{D_j, j = 1, \ldots, N_{\text{вых}}\}_m$, $m = 1, \ldots, M$. Требуется подобрать значения коэффициентов связи $J_{ji}^{(n)}$, $i = 1, \ldots, N_n$, $j = 1, \ldots, N_{n+1}$, $n = 1, \ldots, N$, и, вообще говоря, число промежуточных слоев так, чтобы при подаче на вход системы любого изображения из заданного набора на выходе появлялась соответствующая ему картина, или как можно более близкая к ней, обеспечивающая наименьшую ошибку узнавания.

Ошибку узнавания зададим в виде функционала, зависящего от коэффициентов $J_{ji}^{(n)}$:

$$E = \frac{1}{2} \sum_{m=1}^{M} \sum_{j=1}^{N_{\text{вых}}} (D_{j,m} - O_{j,m})^2, \tag{8.74}$$

где $O_{j,m} = y_{j,m}^{(N)}$ — выходной сигнал элементов последнего слоя; для решения задачи обучения будем минимизировать E по всем $J_{ji}^{(n)}$, $i = 1, \ldots, N_n$, $j = 1, \ldots, N_{n+1}$, $n = 1, \ldots, N$.

Для минимизации (8.74) воспользуемся методом наискорейшего спуска, для чего требуется знать градиент E, т.е. вычислить все частные производные E по каждому коэффициенту связи $J_{ji}^{(n)}$. Рассмотрим случай, когда задано единственное эталонное изображение, тогда индекс m, нумерующий изображения обучающей последовательности, можно опустить. Это упрощение не является принципиальным, так как если задана последовательность из нескольких эталонных изображений, то производная $\partial E/\partial J_{ji}^{(n)}$ будет равна сумме всех таких производных, вычисленных для каждого изображения.

Для вычисления $\partial E/\partial J_{ji}^{(n)}$ воспользуемся тождеством

$$\frac{\partial E}{\partial y_i^{(n-1)}} = \sum_j \frac{\partial E}{\partial y_j^{(n)}} \frac{\partial y_j^{(n)}}{\partial x_j^{(n)}} \frac{\partial x_j^{(n)}}{\partial y_j^{(n-1)}},$$

которое с учетом (8.72) и (8.73) можно переписать в виде

$$\frac{\partial E}{\partial y_i^{(n-1)}} = \sum_j \frac{\partial E}{\partial y_j^{(n)}} y_j^{(n)}(1 - y_j^{(n)}) J_{ji}^{(n-1)}.$$

Это равенство позволяет вычислять производную $\partial E/\partial y_i^{(n-1)}$ для каждого слоя элементов сети по известным значениям производной $\partial E/\partial y_i^{(n)}$ на следующем слое. Для выходного слоя эти производные вычисляются из (8.74):

$$\frac{\partial E}{\partial y_i^{(N)}} = y_i^{(N)} - D_i.$$

Таким образом, начиная с последнего слоя и двигаясь в направлении, противоположном распространению сигналов сети, можно вычислить все производные $\partial E/\partial y_i^{(n-1)}$, $n = 2, 3, \ldots, N$, $i = 1, \ldots, N_{n-1}$. Для вычисления производной E по $J_{ji}^{(n)}$ воспользуемся формулой

$$\frac{\partial E}{\partial J_{ji}^{(n-1)}} = \frac{\partial E}{\partial y_j^{(n)}} \frac{dy_j^{(n)}}{dx_j^{(n)}} \frac{\partial x_j^{(n)}}{\partial J_{ji}^{(n-1)}},$$

которая с учетом (8.72), (8.73) перепишется в виде

$$\frac{\partial E}{\partial J_{ji}^{(n-1)}} = \frac{\partial E}{\partial y_j^{(n)}} y_j^{(n)}(1 - y_j^{(n)}) y_i^{(n-1)}.$$

Алгоритм минимизации E состоит в изменении каждого коэффициента связи $J_{ji}^{(n)}$ в соответствии с формулой

$$\Delta J_{ji}^{(n)} = -\epsilon \frac{\partial E}{\partial J_{ji}^{(n)}}.$$

Минимизация E может производиться как для каждого эталонного изображения, так и по всем изображениям. В последнем случае в качестве $\partial E/\partial J_{ji}^{(n)}$ используется сумма производных, вычисленных для каждого эталонного изображения. Процесс продолжается, пока не будет достигнута заданная точность распознавания. Если в процессе обучения не удается достичь приемлемых значений ошибки E, следует добавить еще один слой элементов и повторить обучение сначала.

Обученная таким образом система способна узнавать ситуации, характеризующиеся входными сигналами, сходными с эталонными, путем формирования выходных образов, близких к запомненным при обучении, причем понятие “сходности” ситуации при этом априори не формализовано, а выработано системой в процессе обучения. Такой подход весьма перспективен для решения сложных неформализуемых задач. Эксперименты показали перспективность обучающихся нейронных сетей при анализе событий в ядерной физике в стримерных камерах и других детекторах. В качестве примера решения задачи узнавания приведем систему, запоминающую родственные связи группы людей. Машине предъявляли обучающие выборки сигналов, представляющих в закодированном виде информацию о ближайших родственниках: кто является сыном или дочерью каких родителей. После обучения система способна отвечать на вопросы типа “кто является тетей такого-то?..” и другие подобные. Сообщается также о системах, обученных чтению вслух английских текстов, при этом был использован метод обратного распространения ошибок.

8.8.3. Инвариантные сети. На практике нейронные сети нашли широкое применение при анализе и распознавании изображений. Специфика задач анализа изображений трехмерных сцен и объектов состоит в том, что с точки зрения анализа и распознавания объектов часто оказываются несущественными расположение устройства, формирующего изображение, относительно сцены, определяющее ракурс сцены, расположение ее деталей на изображении, освещение сцены и другие условия наблюдения. Методам, инвариантным относительно условий регистрации, посвящен рассмотренный ниже морфологический анализ изображений. Избавится от зависимости методов анализа сцен по их изображениям от геометрии наблюдения объектов в простейшем случае контурных изображений можно путем специального конструирования нейронной сети в виде сетчатки светочувствительных датчиков, соединенных между собой специальным образом так,

что выходные сигналы сетчатки не зависели от расположения изображения на сетчатке, его масштабов, поворотов, зеркальных отражений, растяжений или сжатий. Этими преобразованиями в значительной степени можно моделировать изменение ракурса наблюдаемых объектов сцены.

Общий подход к такому описанию можно сформулировать следующим образом. Пусть каждому объекту или сцене соответствует некоторая функция (изображение) $f(\mathbf{x})$, заданная на поле зрения X. Значение $f(\mathbf{x})$ в точке $\mathbf{x} \in X$ может означать, например, яркость изображения объекта в заданной точке поля зрения. Множеству объектов, подлежащих классификации, соответствует, таким образом, множество L описаний $f(\mathbf{x})$ на поле зрения X — т.е. функциональное пространство.

Пусть G — группа преобразований X, так, что $g\mathbf{x}$ при $g \in G$, $\mathbf{x} \in X$, есть результат преобразования $g \in G$ точки $\mathbf{x} \in X$. Соответственно, описание $f(\mathbf{x})$ деформируется в описание $f^*(\mathbf{x}) = f(g\mathbf{x})$, причем и описание $f(\mathbf{x})$, и деформированное описание $f^*(\mathbf{x})$ при классификации должны отождествляться, так как они соответствуют одному и тому же объекту. При этом предполагается, что если $f(\mathbf{x}) \in L$, то и $f^*(\mathbf{x}) \in L$ при всех $g \in G$, так что L разбивается на непересекающиеся классы эквивалентных описаний:

$$L = \bigcup_f K_f.$$

Каждый класс K_f состоит из G-эквивалентных описаний $f(g\mathbf{x})$, где g пробегает всю группу G. Задача классификации состоит в том, чтобы получить описания классов K_f: класс K_f определяет $f(\mathbf{x})$ с точностью до преобразования $f(\mathbf{x}) \to f(g\mathbf{x})$.

Например, пусть X — плоскость фотографии, и $f(\mathbf{x})$ — ее яркость (плотность почернения фотобумаги). Если G — группа вращений поля зрения, то $f(g\mathbf{x})$ — яркость повернутой фотографии. В задачах классификации фотографий $f(\mathbf{x})$ и $f(g\mathbf{x})$ попадают в один и тот же класс K_f.

Получим для случая контурных изображений набор функционалов, характеризующих $f(\mathbf{x})$ с точностью до афинного преобразования поля зрения.

Рассмотрим контурное изображение, заданное как бинарная функция двух переменных из подмножества X плоскости $\mathcal{R}_2$ — поля зрения:

$$f(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \quad \text{принадлежит контуру} \quad \varphi, \\ 0, & \mathbf{x} \quad \text{не принадлежит контуру} \quad \varphi. \end{cases}$$

Здесь контур φ — упорядоченный набор точек поля зрения X. Будем задавать его набором конечного числа точек $\mathbf{x}_i$, $i = 1, \ldots, N$.

(Выбирая N достаточно большим, можно сколь угодно точно приблизить достаточно гладкую кривую на плоскости.) Введем систему координат на поле зрения X так, чтобы каждая точка $\mathbf{x}_i$ контура φ, $i = 1, \ldots, N$, задавалась парой чисел (x_i, y_i), $i = 1, \ldots, N$. Рассмотрим координатное пространство $\mathcal{R}_{2N}$ точек $\varphi = (x_1, y_1, \ldots, x_N, y_N)$ и группу преобразований G пространства $\mathcal{R}_{2N}$, таких, что при $g \in G$ соотношение $\varphi^* = g(\varphi)$ задает новое положение контура φ на поле зрения X. Под группой G будем понимать такие преобразования, которые получаются путем афинных преобразований каждой точки $\mathbf{x}_i \in X \subset \mathcal{R}_2$, $i = 1, \ldots, N$, контура φ. С точки зрения классификации интерес представляют лишь те свойства контура φ, которые не изменяются при таком преобразовании — т.е. инварианты группы G.

Обозначим $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$, $\mathbf{x}^* = \begin{pmatrix} x^* \\ y^* \end{pmatrix}$, $\mathbf{a} = \begin{pmatrix} a_x \\ a_y \end{pmatrix}$, $\widehat{A} = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}$, причем $\det A \neq 0$, и рассмотрим афинное преобразование g_2 точки $\mathbf{x} \in X$:

$$\mathbf{x}^* = g_2 \mathbf{x} \equiv \widehat{A}\mathbf{x} + \mathbf{a}.$$

Преобразование $g \in G$ получается применением афинного преобразования g_2 к каждой точке $\mathbf{x}_i$, $i = 1, \ldots, N$, контура $\varphi \in \mathcal{R}_{2N}$ и, значит, определяется шестью параметрами.

Для общей афинной группы N точкам на поле зрения X соответствует $2N - 6$ инвариантов J_α, $\widetilde{J}_\alpha$; $\alpha = 1, \ldots, N-3$:

$$J_\alpha = \frac{\Delta_\alpha}{\Delta_0}, \quad \widetilde{J}_\alpha = \frac{\widetilde{\Delta}_\alpha}{\Delta_0},$$

где

$$\Delta_\alpha = \det \begin{pmatrix} \xi_{\alpha+2} & \xi_{\alpha+1} \\ \eta_{\alpha+2} & \eta_{\alpha+1} \end{pmatrix}, \quad \widetilde{\Delta}_\alpha = \det \begin{pmatrix} \xi_{\alpha+2} & \xi_{\alpha} \\ \eta_{\alpha+2} & \eta_{\alpha} \end{pmatrix},$$

а $\xi_\alpha = x_\alpha - x_{\alpha+1}$, $\eta_\alpha = y_\alpha - y_{\alpha+1}$. Каждый инвариант J_α, $\widetilde{J}_\alpha$, зависит от координат четырех точек контура φ, и по определению численно не меняется при афинных преобразованиях поля зрения. Таким образом, набор инвариантов определяет описание контурного изображения, не зависящее от указанных преобразований поля зрения.

Заметим, что любые три точки поля зрения, не лежащие на одной прямой, под действием преобразования g могут быть переведены в любые три точки, также не лежаще на одной прямой.

Нейронная сеть, выходные сигналы которой не зависят от положения контурного изображения на поле зрения, должна вычислять указанные инварианты. Дальнейшая классификация может осуществляться методами, рассмотренными выше.

8.8.4. Морфологический анализ изображений. Одной из фундаментальных проблем анализа изображений является создание

математического описания, передающего их содержание, смысл. Описание должно отражать лишь существенные с точки зрения решаемой задачи особенности изображений и не зависеть от несущественных. В морфологическом анализе несущественными характеристиками являются условия регистрации изображений объекта или сцены и параметры регистрирующей аппаратуры. Методы морфологического анализа являются, таким образом, шагом на пути решения проблемы описания семантики изображений.

Хорошо известно, что изображения одной и той же сцены, отвечающие различным условиям освещения и (или) измененным оптическим свойствам объектов могут различаться радикально. Это обстоятельство порождает значительные трудности в прикладных задачах анализа и интерпретации изображений реальных сцен, в которых решение не должно зависеть от условий регистрации изображений. Речь идет, например, о задачах выделения неизвестного объекта на фоне известной местности, известного объекта на произвольном фоне при неконтролируемых условиях освещения, о задаче совмещения изображений одной и той же сцены, полученных в различный спектральных диапазонах и т.д.

Для решения перечисленных задач описываемые здесь методы морфологического анализа изображений оказываются достаточно эффективными.

Основную идею этих методов поясним на следующем примере. Рассмотрим изображение некоторой сцены, сформированное с помощью видеокамеры в видимом диапазоне спектра. Как правило, оно несет достаточно подробную информацию о геометрической форме изображаемых объектов сцены, так как области изображения, все точки которых имеют одинаковые яркости, отображают части поверхности объектов, обладающие одинаковыми геометрическими и оптическими свойствами. Предположим, что эти свойства сохраняются, если изменить условия освещения, диапазон и природу регистрируемого излучения, тогда любые изображения данной сцены, сформированные тем или иным способом, будучи связанными с геометрическими свойствами объектов сцены, могут быть получены путем преобразования яркости видеоизображения. В этой ситуации различная природа устройств формирования изображения или изменившиеся условия освещения и т.п. могут моделироваться путем всевозможных преобразований яркости исходного видеоизображения. Максимальный инвариант такого класса преобразований определит все то, что относится к данной сцене и не зависит от условий формирования изображений — его можно назвать формой. Поскольку этот инвариант еще не определяет полностью всю форму сцены, его называют формой и з о б р а ж е н и я сцены. Отличия формы одного изображения сцены от формы другого характеризуют свойства сцены,

не связанные с условиями регистрации, и могут отражать изменения в содержании изображаемого — например, наличие новых объектов или отсутствие старых. Яркость инфракрасного изображения, или, например, изображения доплеровского измерителя скорости, несущую информацию качественного характера о физических свойствах объектов — температурном контрасте или относительной скорости приемника и объекта — естественно связать с геометрической формой изображения сцены и, тем самым, "привязать" к конкретным объектам сцены.

Важным классом задач, предшествующих решению задачи сравнения по форме предъявленных изображений, являются задачи совмещения и выделения заданных фрагментов. Например, дано изображение местности, полученное с некоторой "точки зрения", и на нем выделен определенный фрагмент. На другом изображении, полученном при других условиях и (или) в другом диапазоне спектра, требуется указать выделенный фрагмент изображения местности и уточнить "точку зрения", с которой оно получено.

Итак, в основе методов морфологического анализа лежит математическое понятие формы. Под формой изображения понимается максимальный инвариант преобразований изображения, которым оно подвергается при изменении условий наблюдения, изменении параметров регистрирующей аппаратуры и др.

Форма, таким образом, определяется не только исследуемым объектом или сценой, но и тесно связана с моделью регистрации изображения объекта или сцены. Например, если речь идет о задачах сравнения или классификации объектов, состоящих из однородно светорассеивающих плоскостей, по их изображениям, полученным при сохранении ракурса, но при различных условиях освещения, то под формой изображения можно понимать максимальный инвариант преобразования яркостей изображения, каковым является проектор (в некоторых случаях — ортогональный) в гильбертовом пространстве изображений на множество кусочно постоянных изображений. Построение формы изображений является основополагающей частью задач морфологического анализа и рассмотрена в данном разделе.

Форма изображения. Поясним понятие формы изображения сначала на простом примере. Рассмотрим изображения "единицы", представленные на рис. 8.18.

Представленные на рис. 8.18,*а*–*в*, *д* изображения обладают одной и той же формой (в обычном понимании) и отличаются лишь условиями регистрации. Форма изображения ровного поля зрения рис. 8.18,*г* более проста, чем форма изображений рис. 8.18,*а*–*в*, *д*. Каждое изображение рис. 8.18 зададим как функцию

$$f(x) = c_1\chi_1(x) + c_2\chi_2(x), \quad x \in \mathbf{X}, \tag{8.75}$$

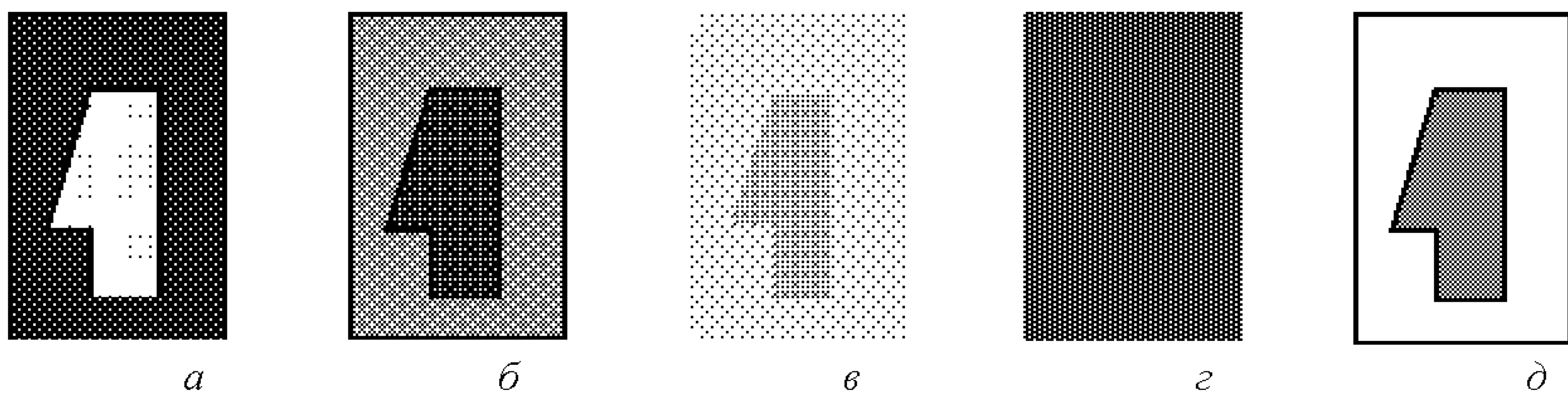


Рис. 8.18. Изображения *а, б, в, д* эквивалентны по форме, изображение *г* не сложнее, чем изображения *а, б, в, д,* и не эквивалентно им.

определенную на поле зрения $\mathbf{X}$ и принимающую два значения: c_1 и c_2 на его подмножествах A_1 — "единице" и A_2 — "фоне" соответственно. В соотношении (8.75)

$$\chi_i(x) = \begin{cases} 1, & x \in A_i, \\ 0, & x \notin A_i, \end{cases} \quad i = 1, 2,$$

— индикаторная функция множества A_i, $i = 1, 2$, причем множества A_1 и A_2 не пересекаются, $A_1 \cap A_2 = \emptyset$, и их объединение $A_1 \cup A_2 = \mathbf{X}$. В более простом по форме изображении рис. 8.18,*г* яркости множеств A_1 и A_2 совпадают, и изображение рис. 8.18,*г* может быть получено из любого из изображений рис. 8.18,*а–в*, *д* путем преобразования их яркости, так, что что результат преобразования зависит только от яркости преобразуемой точки изображения и не зависит от ее расположения на поля зрения. Аналогичными преобразованиями яркости изображения рис. 8.18,*г* можно получить лишь изображения ровного поля зрения, эквивалентные по форме изображениям рис. 8.18,*г*, и нельзя получить изображения более сложной формы.

Изображение "единицы" (8.75) удобно рассматривать как элемент класса всех возможных изображений на поле зрения $\mathbf{X}$, под которым будем понимать лебеговский класс $\mathcal{L}_2 = \mathcal{L}_2(\mathbf{X})$ измеримых и интегрируемых с квадратом функций на $\mathbf{X}$. Поскольку $\mathcal{L}_2$ — евклидово пространство относительно скалярного произведения

$$(f, g) = \int_{\mathbf{X}} f(x) g(x) dx, \quad f(\cdot),\ g(\cdot) \in \mathcal{L}_2,$$

то мы можем использовать наглядную геометрическую терминологию. Например, если в (8.75) допустимы любые значения c_1 и c_2, то множество всех изображений "единицы" будет двумерной плоскостью V_f в $\mathcal{L}_2$:

$$V_f = \{c_1 \chi_1(\cdot) + c_2 \chi_2(\cdot), \quad -\infty < c_1,\, c_2 < \infty\}, \tag{8.76}$$

которая и есть искомый максимальный инвариант, не изменяющийся при произвольных преобразованиях c_1 и c_2 и определяющий изображение с точностью до такого преобразования. Его можно задать также

равенством
$$V_f = \{F(f(\cdot)), \quad F \in \mathbf{F}\}, \tag{8.77}$$
в котором функция $f(\cdot)$ определена равенством (8.75) при $c_1 \neq c_2$, а $\mathbf{F}$ — класс всех (борелевских) функций, определенных на действительной прямой и принимающих числовые значения, моделирующий различные условия регистрации изображений "единицы". Если в (8.75) $c_1 > c_2$, и среди "единиц" не может быть "негативных изображений", то вместо (8.76) следует писать
$$V_f = \{c_1\chi_1(\cdot) + c_2\chi_2(\cdot), \quad -\infty < c_1 \le c_2 < \infty\}, \tag{8.78}$$
а в (8.77) $\mathbf{F}$ — класс монотонно неубывающих функций. Множество V_f теперь геометрически изображается полуплоскостью (конусом) в $\mathcal{L}_2$.

Следует отметить, что, кроме изображений "единицы", V_f содержит изображение однородного поля зрения, у которого $c_1 = c_2$. Форму таких изображений естественно считать более простой, чем форма "единицы" — оно содержит лишь одну область одинаковой яркости. Изображение "ровного поля зрения" $f_0 = c_0 = \mathrm{const}$ можно получить из любого изображения вида (8.75) преобразованием его яркостей с помощью функции $F(\cdot)$ такой, что $F(c_1) = F(c_2) = c_0$. Распространяя этот пример на произвольные изображения, будем считать, что $\widetilde{f}(\cdot)$ по форме не сложнее, чем $f(\cdot)$, если найдется такая функция $F(\cdot) \in \mathbf{F}$, что $\widetilde{f}(\cdot) = F(f(\cdot))$. Будем отмечать этот факт, записывая $\widetilde{f} \prec f$. Про такие изображения будем говорить также, что они сравнимы по форме с f. Соответственно, множество V_f (8.76), (8.77), состоит из изображений, сравнимых по форме с f, причем форма каждого из них не сложнее, чем форма f. Понятно, что сказанное переносится и на произвольные изображения.

Определение. Пусть $\mathbf{F}$ — класс функций, определенных и принимающих значения на действительной прямой, $\mathbf{F}_f$ — подкласс $\mathbf{F}$, выделенный условием $\mathbf{F}_f = \{F(f(\cdot)), \quad F \in \mathbf{F},\ F(f(\cdot)) \in \mathcal{L}_2\}$. *Формой изображения* $f(\cdot)$ назовем подмножество
$$V_f = \{F(f(\cdot)), \quad F \in \mathbf{F}_f\}$$
гильбертова пространства $\mathcal{L}_2$. Будем говорить, что форма изображения f' не сложнее, чем форма f, и писать $f' \prec f$, если $f' \in V_f$. Изображения f' и f назовем эквивалентными по форме, если $f' \prec f$ и $f \prec f'$. Факт эквивалентности изображений будем отмечать как $f \sim f'$. Изображения f' и f назовем совпадающими по форме, если $V_{f'} = V_f$, в этом случае будем писать $f \approx f'$.

Согласно определению, форма V_f изображения f состоит из изображений $f' \prec f$, следовательно, все изображения из V_f сравнимы по форме с f, но не обязательно сравнимы по форме между собой. Что

касается других свойств множества $\mathbf{F}$, определяющего модель формирования яркости изображений, то в большинстве случаев его можно считать выпуклым. При этом выпуклым будет и V_f, поскольку если $f_1 \prec f$ и $f_2 \prec f$, то и $af_1 + bf_2 \prec f$ для любых $a > 0$, $b > 0$, $a + b = 1$. Далее множество $\mathbf{F}$ предполагается выпуклым, а V_f — замкнутым в $\mathcal{L}_2$. Выпуклость и замкнутость V_f гарантирует существование оператора P_f проецирования на V_f, который определяется с помощью решения следующей задачи наилучшего приближения в $\mathcal{L}_2$:

$$||P_f g - g||^2 = \inf \left\{ ||f' - g||^2 \mid f' \in V_f \right\} \quad \text{для любого } g \in \mathcal{L}_2. \qquad (8.79)$$

Поскольку $P_f g \in V_f$, изображение $P_f g$ является самым лучшим приближением изображения g изображениями, форма которых не сложнее, чем форма f; соответственно, изображение $g - P_f g$ представляет все то, что отличает g по форме от f. Поскольку V_f замкнуто, то $V_f = \{g \in \mathcal{L}_2, \ \ P_f g = g\}$, т.е. оператор P_f вполне определяет форму V_f, и его также можно назвать формой изображения f. Отношение $\approx$ транзитивно: если $f'' \approx f'$ и $f' \approx f$, то $f'' \approx f$. Что касается отношений $\prec$ и $\sim$, то для них свойство транзитивности в общем случае не выполняется. Оно будет выполняться, если множество $\mathbf{F}$ наделить свойствами полугруппы относительно естественной операции композиции функций, а именно, если из включений $F(\cdot) \in \mathbf{F}$, $F_1(\cdot) \in \mathbf{F}$ будет следовать включение $F(F_1(\cdot)) \in \mathbf{F}$. В этом случае $\sim$ и $\approx$ означают одно и то же.

В заключение этого раздела дадим еще одно определение.

Определение. *Формой в широком смысле* изображения f назовем оператор ортогонального в $\mathcal{L}_2$ проецирования на минимальное линейное подпространство $\mathcal{L}_2$, содержащее V_f.

Построение формы изображения. Как уже говорилось выше, понятие формы является центральным в морфологическом анализе; построением формы начинается решение задачи морфологической классификации, обнаружения, выделения отличий по форме и др.

Как следует из предыдущего пункта, форма кусочно постоянного изображения будет задана, если определены подмножества поля зрения $\mathbf{X}$, обладающие одинаковыми яркостями, и указан класс изменения этих яркостей. Один из способов задания таких областей состоит в определении фрагментов поля зрения, изображающих однородно светящиеся или отражающие грани объекта: приписывая им всевозможные яркости, получим форму изображения объекта как множество его возможных изображений. Если же столь подробные сведения об объекте исследования не доступны, можно построить форму V_f по какому-либо одному изображению f, указав на нем множества одинаковой яркости и задав пределы изменения яркостей, соответствующие изменяющимся условиям наблюдения.

Пример. Пусть

$$f(x) = \sum_{i=1}^{N} c_i \chi_i(x),$$

$\chi_i(\cdot)$ — индикаторные функции множеств A_i, $A_i \cap A_j = \emptyset$, $\bigcup\limits_{i=1}^{N} A_i = X$, и $\mathcal{F}$ — класс всех (борелевских) функций, определенных на $\mathcal{R}_1$ со значениями в $\mathcal{R}_1$. Тогда

$$V_f = \left\{ \sum_{i=1}^{N} c_i \chi_i(x), \quad -\infty, c_i < \infty, \quad i = 1, \ldots, N \right\}$$

— форма изображения $f(\cdot)$. Проектор P_f на эту форму как решение задачи (8.79) дается формулой: для любого $g(\cdot) \in \mathcal{L}_2$ выполнено

$$P_f g(x) = \sum_{i=1}^{N} \frac{(\chi_i, g)}{\|\chi_i\|^2} \chi_i(x) = \sum_{i=1}^{N} \frac{\int\limits_X \chi_i(z) g(z) dz}{\int\limits_X \chi_i^2(z) dz} \chi_i(x).$$

Выделение отличий изображений по форме. Пусть форма в широком смысле изображения f задана в виде оператора ортогонального проецирования P_f в $\mathcal{L}_2$ на линейное подпространство $\mathcal{R}(P_f) \subset \mathcal{L}_2$, предъявляется априори произвольное изображение $g \in \mathcal{L}_2$ и требуется указать, чем отличается по форме изображения f и g. Как отмечено выше, изображение $g - P_f g$ представляет все то, что отличает g по форме от любого $f \in \mathcal{R}(P_f)$.

Обнаружение объекта по его изображению. Пусть задана форма в широком смысле P_f изображения объекта, предъявлено изображение g и требуется высказаться, является ли g изображением этого объекта. Морфологический метод решения этой задачи состоит в проверке равенства $P_f g = g$; если оно выполнено, то изображение g может быть порождено заданным объектом при некоторых условиях регистрации.

На практике изображение $f(\cdot)$ задается в виде матрицы яркостей размера $N_x \times N_y$, $N_x N_y = M$, каждый матричный элемент f_{ij} представляет собой яркость точки $(x_i, y_j) \in \mathcal{R}_2$, $i = 1, \ldots, N_x$, $j = 1, \ldots, N_y$. В этом случае удобно рассматривать изображение как элемент конечномерного евклидова пространства R_M. Если предъявленное изображение содержит шум, т.е. анализу доступно лишь изображение $\xi = g + \nu$ и $\nu \in \mathcal{R}$ контролируется нормальным распределением $\mathcal{N}(0, \sigma^2 I)$, то задача обнаружения объекта ставится как задача проверки статистической гипотезы о параметре нормального распределения случайного вектора $\xi \in \mathcal{R}$. Гипотеза состоит в том, что $\xi \sim \mathcal{N}(P_f g, \sigma^2 I)$, где $g \in \mathcal{R}_M$ — некоторый вектор, против

альтернативы $\xi \sim \mathcal{N}(z, \sigma^2 I)$, где $z \neq P_f z$. Эта задача инвариантна относительно преобразований пространства наблюдений $\mathcal{R}_M$, состоящих в сдвиге вектора ξ вдоль пространства $\mathcal{R}(P_f)$ значений проектора P_f и вращении и (или) отражении ортогонального дополнения в $\mathcal{R}_M$ к пространству $\mathcal{R}(P_f)$, и соответствующей группы преобразований параметров распределения. Максимальным инвариантом этой группы преобразований является функция $\psi(\cdot) = \|(I - P_f)\cdot\|$, причем статистика $\tau_N(\xi) = (1/\sigma^2)\|(I - P_f)\xi\|^2$ при верной гипотезе имеет распределение χ^2 с $(M - N_f)$ степенями свободы, а при верной альтернативе — нецентральное распределение χ^2 с $(M - N_f)$ степенями свободы и параметром нецентральности, равным $\vartheta^2 = (1/\sigma^2)\|(I - P_f)z\|^2$; здесь N_f — размерность пространства значений проектора P_f. В задаче проверки гипотезы о равенстве нулю параметра нецентральности существует следующий равномерно наиболее мощный критерий:

гипотеза $\vartheta^2 = 0$ принимается тогда и только тогда, когда

$$\frac{1}{\sigma^2}\|(I - P_f)z\|^2 \leq \varepsilon,$$

где ε — корень уравнения

$$\int_0^{\varepsilon} p_{M-N_f}(x, 0)dx = \alpha,$$

а $p_k(x, \vartheta^2)$ — плотность χ^2-распределения с k степенями свободы и параметром нецентральности ϑ^2. В этом и только в этом случае будем считать, что изображение ξ имеет форму, неотличимую в шуме от формы f.

Глава 9

Системы со случайными шумами

9.1. Роль флуктуаций

Ранее уже рассматривалась конструктивная роль случайных флуктуаций параметров системы, приводящая к сложному и необратимому характеру ее эволюции. Эти флуктуации часто связываются с невозможностью абсолютно точного определения фазовых координат системы, что приводит к необходимости поиска иных принципов моделирования, чем описание систем в виде детерминированных дифференциальных или иных эволюционных уравнений. В частности, благодаря флуктуациям фазовых координат разрушаются детерминированные фазовые траектории, исчезают особые точки (стационарные, седловые и т.п.), система как бы "прощупывает" фазовое пространство и блуждает вокруг состояния, предсказываемого решением детерминированного эволюционного уравнения.

Рассмотрим, к каким качественно новым эффектам приводит введение случайных флуктуаций в эволюцию системы. Запишем эволюционное уравнение

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) \tag{9.1}$$

с начальными условиями

$$\mathbf{x}(0) = \mathbf{x}_0. \tag{9.2}$$

Если флуктуируют только начальные данные $\mathbf{x}_0$ и система не является перемешивающей, то неопределенность задания $\mathbf{x}_0$, обусловленная возможными флуктуациями, переносится вдоль фазовых траекторий, не расплываясь по всему фазовому пространству, т.е. неопределенность в начальных условиях трансформируется в неопределенность значений фазовых координат в каждый последующий момент

времени, превращая траекторию в “диффузную трубку” наподобие следа от реактивного самолета. Для систем с перемешиванием эта неопределенность со временем растет и заполняет всю доступную часть фазового пространства.

Если флуктуируют не только начальные данные (9.2), но и правая часть (9.1), то неопределенность фазовых координат растет со временем как для систем с перемешиванием, так и для неперемешивающих систем.

Примером системы с флуктуирующей правой частью (9.1) является броуновское движение. Траектория броуновской частицы описывают случайным процессом Винера $\xi(t)$; дисперсия координаты частицы $\mathbf{E}(\xi(t) - \mathbf{E}\xi(t))^2$, описывающая неопределенность положения частицы в момент времени t, линейно растет со временем.

Если система, описываемая детерминированным уравнением (9.1), имеет аттрактор в виде фокуса $\mathbf{x}_0$, то система без флуктуаций со временем стремится к состоянию $\mathbf{x}_0$, флуктуации же правых частей (9.1) приводят к тому, что сходимость $\mathbf{x}(t)$ к $\mathbf{x}_0$ исчезает, можно лишь говорить о вероятности того, что система с течением времени будет находиться в той или иной окрестности точки $\mathbf{x}_0$.

9.2. Случайные процессы

Для описания динамических систем со случайными флуктуациями воспользуемся аппаратом теории случайных процессов, которые будут служить моделями фазовых траекторий таких систем. Пусть $(\Omega, \mathcal{F}, P)$ — вероятностное пространство, T — некоторое подмножество числовой оси. Случайным процессом назовем числовую функцию $\xi(\omega, t)$, $\omega \in \Omega$, $t \in T$, такую, что при фиксированном t она является случайной величиной (т.е. $\mathcal{F}$-измеримой функцией на Ω). При фиксированном t функция $\xi(\omega, t)$ задает случайную величину, называемую сечением процесса, а при фиксированном ω — функцию аргумента $t \in T$, называемую выборочной функцией, или реализацией. Для того, чтобы описать вероятностные свойства процесса. недостаточно задать лишь распределения

$$F_1(x_1, t_1) = P(\xi(\omega, t_1) < x_1), \quad x_1 \in \mathcal{R}_1,$$

сечения для всех t_1 из множества T, так как эти распределения не дают информации о совместном распределении всех сечений. Более полная информация содержится в совместных распределениях случайных величин $(\xi(t_1), \xi(t_2))$, $(\xi(t_1), \xi(t_2), \xi(t_3)), \ldots$, $(\xi(t_1), \xi(t_2), \ldots$ $\ldots, \xi(t_n)), \ldots$ размерности 2, 3, …, n соответственно:

$$F_2(x_1,t_1;\, x_2,t_2) = P(\xi(\omega,t_1) < x_1,\, \xi(\omega,t_2) < x_2),$$
$$x_1,\ x_2 \in \mathcal{R}_1, \quad t_1,\ t_2 \in T;$$

$$F_3(x_1,t_1;\, x_2,t_2;\, x_3,t_3) = P(\xi(\omega,t_1) < x_1,\, \xi(\omega,t_2) < x_2),\, \xi(\omega,t_3) < x_3),$$
$$x_1,\, x_2,\, x_3 \in \mathcal{R}_1, \quad t_1,\, t_2,\, t_3 \in T;$$

. . . .

$$F_n(x_1,t_1;\ldots;x_n,t_n) = P(\xi(\omega,t_1) < x_1,\ldots,\xi(\omega,t_n) < x_n),$$
$$x_1,\,\ldots,\,x_n \in \mathcal{R}_1, \quad t_1,\,\ldots,\,t_n \in T;$$

. . . .

Функция $F_n(x_1,t_1;\ldots;x_n,t_n)$ называется n-точечной функцией распределения случайного процесса $\xi(t)$, $n = 2,3,\ldots,n,\ldots$

9.2.1. Марковские случайные процессы. Среди всех случайных процессов выделяется важный класс марковских процессов, для которого можно сформулировать значительно более простое описание вероятностных свойств составляющих его процессов, чем задание всех n-точечных функций распределения. Эти процессы еще называют процессами без последействия. Процесс $\xi(t)$ называется марковским, если для любого t_0 при известном значении $\xi(t_0)$ случайные величины $\xi(t)$ при $t > t_0$ не зависят от случайных величин $\xi(s)$ при $s < t_0$. Таким образом, марковские случайные процессы характеризуются тем, что их вероятностные свойства в моменты времени $t > t_0$ полностью определяются состоянием процесса в момент времени t и не зависят от состояний процесса в моменты, предшествующие t_0.

Рассмотрим марковский случайный процесс $\xi(t)$, $t \in T$, такой, что при любом наборе моментов времени $t_1,t_2,\ldots,t_n \in T$ n-мерная случайная величина $(\xi(t_1),\xi(t_2),\ldots,\xi(t_n))$ имеет n-мерную плотность вероятности $p_n(t_1,x_1;t_2,x_2;\ldots;t_n,x_n)$, которая обладает двумя очевидными свойствами.

1. Плотность вероятности $p_n(t_1,x_1;t_2,x_2;\ldots;t_n,x_n)$ инвариантна относительно перестановки любых пар аргументов (t_i,x_i), так как $p_n(t_1,x_1;t_2,x_2;\ldots;t_n,x_n)dx_1dx_2\ldots dx_n$ есть вероятность совместного наступления событий $x_i \le \xi_i(t_i) < x_i+dx_i$, $i = 1,2,\ldots,n$, и не зависит от порядка их перечисления.

2. Все конечномерные плотности p_n для разных n согласованы так, что при $k < n$

$$p_k(t_1,x_1;\ldots;t_k,x_k) =$$

$$= \int\limits_{-\infty}^{\infty} \ldots \int\limits_{-\infty}^{\infty} p_n(t_1,x_1;\ldots;t_k,x_k;t_{k+1},x_{k+1};\ldots;t_n,x_n)dx_{k+1}\ldots dx_n.$$

Согласно определению плотности условной вероятности,

$$p_n(t_1, x_1; \ldots; t_n, x_n) = \\ = p_{n-1}(t_1, x_1; \ldots; t_{n-1}, x_{n-1}) q(t_n, x_n \,|\, t_1, x_1; \ldots; t_{n-1}, x_{n-1}), \tag{9.3}$$

а так как процесс является марковским, то

$$q(t_n, x_n \,|\, t_1, x_1; \ldots; t_{n-1}, x_{n-1}) = q(t_n, x_n \,|\, t_{n-1}, x_{n-1}). \tag{9.4}$$

Условные плотности $q(t_n, x_n \,|\, t_{n-1}, x_{n-1})$ называются переходными плотностями вероятности. Таким образом, из (9.3) и (9.4) получим формулу

$$p_n(t_1, x_1; \ldots; t_n, x_n) = p_{n-1}(t_1, x_1; \ldots; t_{n-1}, x_{n-1}) q(t_n, x_n \,|\, t_{n-1}, x_{n-1}),$$

применяя которую $(n-1)$ раз, прийдем к соотношению

$$p_n(t_1, x_1; \ldots; t_n, x_n) = \\ = p_1(t_1, x_1) q(t_2, x_2 \,|\, t_1, x_1) \ldots q(t_n, x_n \,|\, t_{n-1}, x_{n-1}), \tag{9.5}$$

из которого следует, что для задания n-мерной плотности вероятности марковского процесса достаточно задать лишь две функции — начальную плотность $p_1(t, x)$ и переходную плотность $q(t, x \,|\, \tau, y)$.

9.2.2. Уравнение Смолуховского. Основным в теории марковских процессов является уравнение Смолуховского (оно также носит название уравнения Колмогорова–Чепмена); это уравнение связывает значения переходных плотностей $q(\cdot, \cdot \,|\, \cdot, \cdot)$ в любые три момента времени $t_0 < \tau < t$, t_0, τ, $t \in T$:

$$q(t, x \,|\, t_0, x_0) = \int\limits_{-\infty}^{\infty} q(t, x \,|\, \tau, y) q(\tau, y \,|\, t_0, x_0) dy. \tag{9.6}$$

Действительно, из (9.5) следует

$$p_3(t_0, x_0; \tau, y; t, x) = p_1(t_0, x_0) q(\tau, y \,|\, t_0, x_0) q(t, x \,|\, \tau, y), \\ p_2(t_0, x_0; t, x) = p_1(t_0, x_0) q(t, x \,|\, t_0, x_0), \tag{9.7}$$

$$p_2(t_0, x_0; t, x) = \int\limits_{-\infty}^{\infty} p_3(t_0, x_0; \tau, y; t, x) dy, \tag{9.8}$$

откуда, подставив в (9.8) соответствующие выражения из (9.7) и сокращая на общий множитель $p_1(t_0, x_0)$, прийдем к уравнению Смолуховского (9.6).

Для однородных по времени марковских процессов переходная плотность $q(t, x \,|\, \tau, y)$ зависит лишь от длительности временного интервала $(t - \tau)$, так, что $q(t, x \,|\, \tau, y) \equiv q(x \,|\, t - \tau, y)$, и уравнение Смолуховского для него записывается в виде

$$q(x \,|\, t - t_0, x_0) = \int\limits_{-\infty}^{\infty} q(x \,|\, t - \tau, y) q(y \,|\, \tau - t_0, x_0) dy.$$

9.3. Уравнение для плотности вероятности

Для некоторых случаев уравнение для переходных плотностей можно получить в виде дифференциального уравнения в частных производных. Наиболее исследованным является уравнение для так называемых диффузионных процессов, описывающее, в частности, и броуновское движение. С формальной точки зрения для диффузионных процессов должны выполняться следующие условия:

1. Средняя скорость изменения состояния системы, находящейся в момент времени t в состоянии y, существует и равна

$$\lim_{\Delta \to 0} \frac{1}{\Delta} \int\limits_{-\infty}^{\infty} (x - y) q(t + \Delta, x \,|\, t, y) dx = A(y, t). \tag{9.9}$$

В частности, для модели броуновского движения интеграл дает среднюю скорость частицы, находящейся в точке y в момент времени t; величина $A(y, t)$ называется коэффициентом сноса.

2. Среднеквадратичный разброс значений состояний x системы, которых она может достигнуть к моменту времени $t + \Delta$, если в момент времени t она находилась в исходном состоянии y, существует и при малых Δ пропорционален Δ:

$$\lim_{\Delta \to 0} \frac{1}{\Delta} \int\limits_{-\infty}^{\infty} (x - y)^2 q(t + \Delta, x \,|\, t, y) dx = B(y, t). \tag{9.10}$$

Для броуновского движения величина $B(y, t)$ называется коэффициентом диффузии и равна дисперсии скорости броуновских частиц.

3. Вероятность больших измерений состояния за малые интервалы времени Δ мала, так, что

$$\lim_{\Delta \to 0} \frac{1}{\Delta} \int\limits_{-\infty}^{\infty} |x - y|^3 q(t + \Delta, x \,|\, t, y) dx = 0. \tag{9.11}$$

В этих условиях и при достаточной гладкости функций $q(t, x \mid t_0, y)$, $A(x, t)$, $B(x, t)$ переходная вероятность $q(t, x \mid t_0, y)$ удовлетворяет уравнению

$$\frac{\partial q(t, x \mid t_0, x_0)}{\partial t} = -\frac{\partial}{\partial x}(A(x, t)q(t, x \mid t_0, x_0)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(B(x, t)q(t, x \mid t_0, x_0)), \tag{9.12}$$

которое является следствием уравнения Смолуховского. Оно известно под названием уравнения Эйнштейна–Фокера–Планка, его решение должно удовлетворять следующим дополнительным условиям:

$$q(t, x \mid t_0, x_0) \geq 0, \quad \int_{-\infty}^{\infty} q(t, x \mid t_0, x_0)dx = 1, \quad \int_{-\infty}^{\infty} q(t, x \mid t_0, x_0)dx_0 = 1,$$

и начальному условию

$$q(t, x \mid t_0, x_0)\Big|_{t=t_0} = \delta(x - x_0).$$

Если в начальный момент времени t_0 задана плотность распределения $p(t_0, x_0)$ случайной величины $\xi(t_0)$, то двумерная плотность вероятности для любого момента времени t равна

$$p_2(t, x; t_0, x_0) = p(t_0, x_0)q(t, x \mid t_0, x_0),$$

а одномерная плотность случайного процесса $\xi(t_0)$ в произвольный момент времени t равна

$$p_1(t, x) = \int_{-\infty}^{\infty} p(t_0, x_0)q(t, x \mid t_0, x_0)dx_0.$$

Умножая уравнение Эйнштейна–Фокера–Планка (9.12) на $p(t_0, x_0)$ и интегрируя по x_0 по всей числовой оси, получим, что плотность $p_1(t, x)$, так же, как и переходная вероятность $q(t, x \mid t_0, y)$, удовлетворяет тому же уравнению Эйнштейна–Фокера–Планка, т.е.

$$\frac{\partial p_1(t, x)}{\partial t} = -\frac{\partial}{\partial x}(A(x, t)p_1(t, x)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(B(x, t)p_1(t, x)), \tag{9.13}$$

причем его решение должно удовлетворять дополнительным условиям

$$p_1(t, x) \geq 0, \quad \int_{-\infty}^{\infty} p_1(t, x)dx = 1, \quad p_1(t, x)\Big|_{t=t_0} = p(t_0, x_0).$$

9.4. Физические системы с шумами

9.4.1. Уравнение Ланжевена. В физических задачах уравнение (9.13) возникает, например, в следующих ситуациях. Рассмотрим частицу, движущуюся в среде с большой вязкостью так, что ее эволюция описывается дифференциальным уравнением Ланжевена:

$$\dot{x}(t) = K(x(t)) + F(t). \tag{9.14}$$

Это так называемое передемпфированное движение. Здесь $K(x(t))$ задает регулярно действующую силу, а $F(t)$ — случайная составляющая внешней силы, возникающая из-за столкновений выделенной рассматриваемой частицы с другими частицами среды. При выполнении предположений (9.9)–(9.11) плотность вероятности $p(t, x)$ найти частицу в точке x в момент времени t дается уравнением Эйнштейна–Фокера–Планка

$$\dot{p} = -\frac{\partial}{\partial x}(Kp) + \frac{1}{2}Q\frac{\partial^2 p}{\partial x^2}, \tag{9.15}$$

где Q — среднее значение квадрата силы $F(t)$ (не зависящее от x). Это уравнение может быть получено из (9.13) в предположении, что случайное блуждание, описываемое эволюционным уравнением (9.14), является однородным по времени марковским процессом, при этом $q(t, x \,|\, t_0, x_0) = q(x \,|\, t - t_0, x_0)$, при этом коэффициент $A(x, t)$ в уравнениях (9.13) и (9.12) равен $K(x)$ и не зависит от времени, а $B(x, t) = Q$ и не зависит ни от времени, ни от координаты.

Рассмотрим сначала стационарные решения уравнения (9.15), для чего перепишем его в виде

$$\dot{p} + \frac{d}{dx}\left(Kp - \frac{1}{2}Q\frac{dp}{dx}\right) = 0, \tag{9.16}$$

и введем обозначение $j = Kp - (Q/2)(dp/dx)$. Величину j назовем потоком вероятности. В терминах плотности p и потока j уравнение (9.16) перепишется в виде

$$\dot{p} + \operatorname{div} j = 0,$$

которое можно интерпретировать как уравнение непрерывности: изменение плотности вероятности равно дивергенции потока вероятности, взятой с обратным знаком.

В многомерном случае, когда состояние системы описывается вектором $\mathbf{x}(t)$, уравнение Эйнштейна–Фокера–Планка имеет вид

$$\frac{\partial p}{\partial t} = -\operatorname{div}(\mathcal{K}p) + \frac{1}{2}\sum_{ij} Q_{ij}\frac{\partial^2}{\partial x_i \partial x_j}p,$$

где $\operatorname{div}(\mathcal{K}p) = \sum_{i=1}^{n} (\partial K_i(\mathbf{x})p(\mathbf{x},t))/(\partial x_i)$. Аналогично одномерному случаю, введем поток вероятности $\jmath = (j_1, \ldots, j_n)$, где

$$j_k = K_k p - \frac{1}{2}\sum_{l=1}^{n} Q_{kl}\frac{\partial p}{\partial x_l},$$

и запишем n-мерное уравнение Эйнштейна–Фокера–Планка как уравнение непрерывности

$$\dot{p} + \operatorname{div}\jmath = 0.$$

Стационарные решения этих уравнений определяются условием $\dot{p} = 0$, то есть $\jmath = \mathrm{const}$. В одномерном случае для получения стационарного решения уравнения (9.16) достаточно учесть граничные условия, обращающие плотность p в нуль на бесконечности. Они приводят к тому, что $j \to 0$ при $x \to \pm\infty$, и в сочетании с условием $j = \mathrm{const}$ означают, что поток вероятности равен нулю на всей числовой оси $x \in \mathcal{R}_1$. Отсюда

$$\frac{1}{2}Q\frac{dp}{dx} = Kp \tag{9.17}$$

и

$$p(x) = N\exp\left(-\frac{2V(x)}{Q}\right), \quad \text{где} \quad V(x) = -\int_{x_0}^{x} K(z)dz.$$

Величине $V(x)$ можно придать смысл потенциала детерминированного силового поля, действующего на частицу согласно уравнению Ланжевена (9.14). Константа N определяется условием нормировки $\int_{-\infty}^{\infty} p(x)dx = 1$. В многомерном случае для стационарного решения имеем $\operatorname{div}\jmath = 0$, откуда еще не следуют уравнения типа (9.17). Для простоты положим, что выполнены так называемые условия потенциальности

$$K_k = -\frac{\partial}{\partial x_k}V(\mathbf{x}).$$

После приведения матрицы Q к диагональной форме прийдем к системе независимых уравнений, аналогичных (9.17).

9.4.2. Движение в потенциальном поле. Рассмотрим примеры. Пусть $K(x) = -\alpha x$, тогда $V(x) = \alpha x^2/2$, график потенциала приведен на рис. 9.1. На этом же рисунке пунктиром изображен график плотности вероятности, полученного при решении стационарного уравнения Эйнштейна–Фокера–Планка в виде

$$p(x) = \frac{\alpha}{\sqrt{2\pi}}e^{-\alpha x^2}.$$

Передемпфированное движение частицы в силовом поле с потенциалом, изображенным на рис. 9.1, привело бы к тому, что с течением времени частица стремилась бы к точке $x = 0$ — стационарной устойчивой точке. Однако наличие случайных сил в уравнении (9.14), описывающем движение частицы, приводит к тому, что стационарная точка "размывается", случайные силы все время "подталкивают" частицу, и можно говорить лишь о вероятности найти частицу в окрестности той или иной точки.

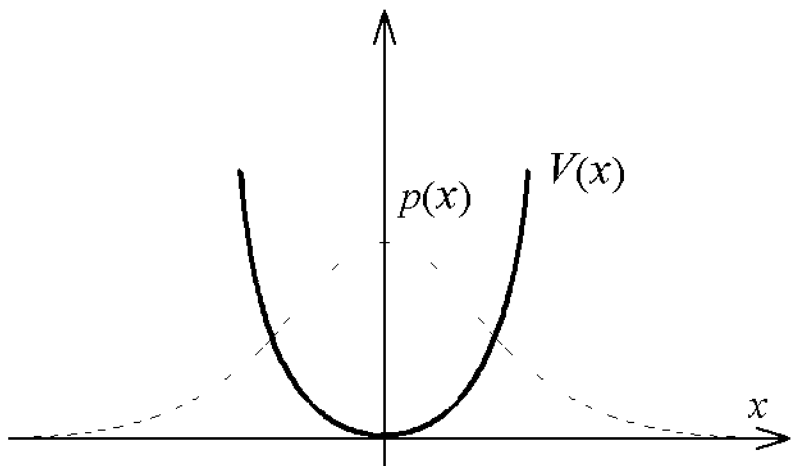


Рис. 9.1. Потенциал силового поля и вероятность найти частицу в окрестности точки равновесия

Если $K(x) = -\alpha x - \gamma x^2 - \beta x^3$, то $V(x) = \alpha x^2/2 + \gamma x^3/3 + \beta x^4/4$. Если $\gamma = 0, \beta > 0$, то при изменении α от положительных значений к отрицательным в детерминированной системе происходит бифуркация, см. рис. 9.2, что отражается на системе со случайными воздействиями в

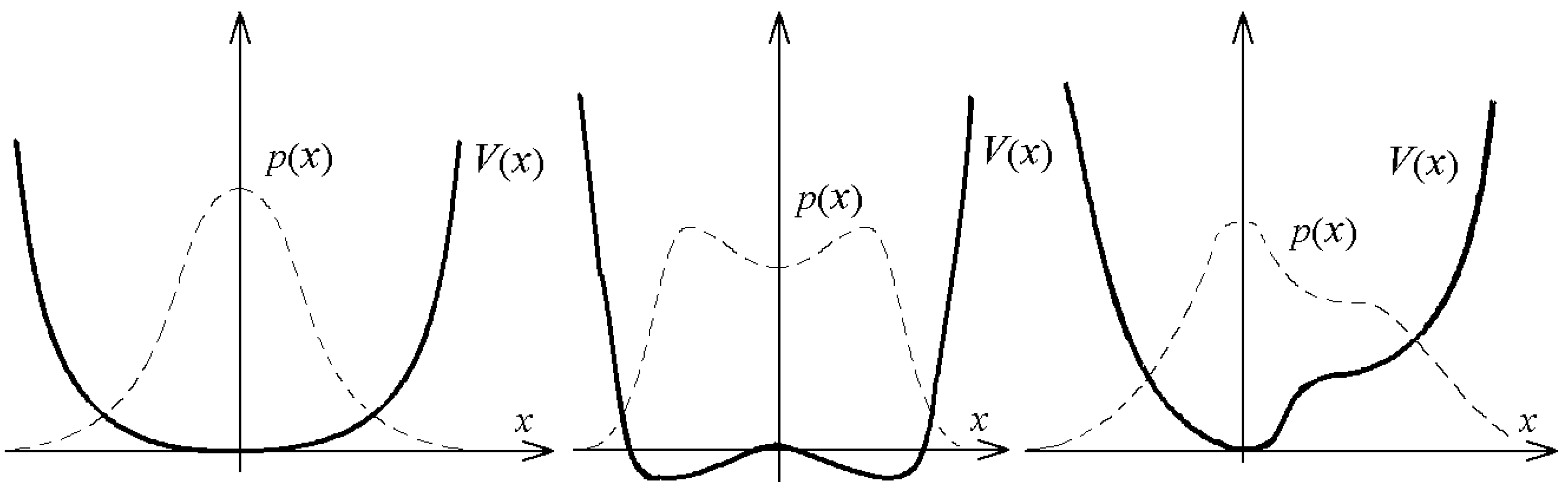


Рис. 9.2. Бифуркация в системах со случайными шумами

переходе от унимодальной плотности к двумодальной. При $\gamma \neq 0$ происходим потеря симметрии распределения.

9.4.3. Барометрическая формула. Рассмотрим теперь движение частицы в однородном силовом поле (например, поле тяжести Земли) с отражающей границей при $x = 0$, потенциал такого поля изображен на рис. 9.3.

В этом случае стационарное решение

$$p(x) = Ce^{-2Kx/Q}, \tag{9.18}$$

и если силовое поле — это поле тяжести Земли, а случайные силы определяются "тепловым движением", так, что $K = mg$, $Q = 2kT$, где g — ускорение свободного падения, k — постоянная Больцмана, T — абсолютная температура, то (9.18) дает известную барометрическую формулу.

9.4.4. Нестационарные решения уравнения Эйнштейна–Фокера–Планка. Рассмотрим, наконец, нестационарное решение уравнения (9.12), для чего положим, что марковский процесс однороден в пространстве так, что $A(x,t)$ и $B(x,t)$ не зависят от времени, и уравнение (9.12) приобретает вид

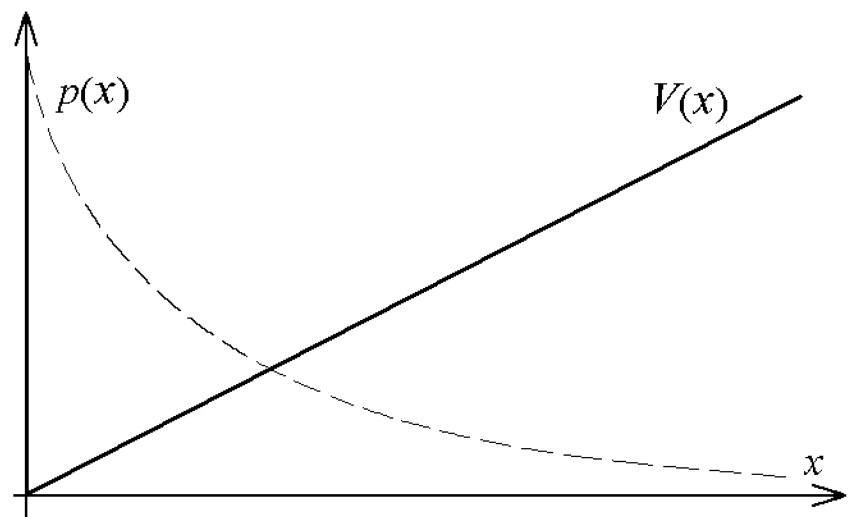


Рис. 9.3. Потенциал поля тяжести и плотность вероятности

$$\frac{\partial q(t, x-x_0 \,|\, t_0)}{\partial t} = -A(t)\frac{\partial q(t, x-x_0 \,|\, t_0)}{\partial x} + \frac{B(t)}{2}\frac{\partial^2 q(t, x-x_0 \,|\, t_0)}{\partial x^2}.$$

Заменим переменные (x,t) на переменные (y,τ) по формулам

$$y = x - x_0 - \int_{t_0}^{t} A(s)ds, \quad \tau = \int_{t_0}^{t} B(s)ds,$$

и прийдем в результате к уравнению теплопроводности

$$\frac{\partial q}{\partial \tau} = \frac{1}{2}\frac{\partial^2 q}{\partial y^2},$$

решение которого хорошо известно. В частности, если $x \in \mathcal{R}_1$ и начальное условие задано в виде δ-функции, возвращаясь к исходным переменным, имеем

$$q(t, x-x_0 \,|\, t_0) = \left(2\pi\int_{t_0}^{t} B(s)ds\right)^{-1/2} \exp\left\{-\frac{1}{2}\frac{\left(x - x_0 - \int_{t_0}^{t} A(s)ds\right)^2}{\int_{t_0}^{t} B(s)ds}\right\},$$

и при $A(t) = 0$, $B(t) = \text{const}$, имеем

$$q(t, x-x_0 \,|\, t_0) = \frac{1}{(2\pi B(t-t_0))^{1/2}} exp\left(-\frac{(x-x_0)^2}{2B(t-t_0)}\right),$$

т.е. в среднем смещение частицы равно нулю, а дисперсия (средний квадрат) смещения пропорциональна времени эволюции — такое поведение свойственно броуновской частице.

9.5. Теория второго порядка

Так же, как классический математический анализ основан на понятии предела функции, можно построить анализ случайных процессов, если в основу его положить понятие предела в среднем квадратичном (с.к.). Речь в этом разделе пойдет о комплексных случайных процессах, у которых существуют математическое ожидание и дисперсия в каждый момент времени.

9.5.1. Сходимость в среднем квадратичном. Рассмотрим последовательность случайных величин $\{\xi_k\}$. Будем говорить, что $\{\xi_k\}$ сходится в среднем квадратичном (с.к.) к случайной величине ξ при $k \to \infty$, если $\mathbf{E}|\xi_k - \xi|^2 \to 0$ при $k \to \infty$. Этот факт обозначим следующим образом:

$$\xi = \underset{k\to\infty}{\text{l.i.m.}}\, \xi_k.$$

Для сходимости в среднем квадратичном необходимо и достаточно, чтобы $\mathbf{E}|\xi_k - \xi_n|^2 \to 0$ при $k, n \to \infty$.

Можно указать и другой критерий: для сходимости последовательности $\{\xi_k\}$ в с.к. необходимо и достаточно, чтобы числовая последовательность $\mathbf{E}\xi_k\overline{\xi_n}$ имела предел при независимом стремлении k и n к бесконечности. Здесь черта над символом ξ означает комплексное сопряжение, при этом имеется в виду, что рассматриваются комплексные случайные величины $\xi = \operatorname{Re}\xi + i\operatorname{Im}\xi$, где $\operatorname{Re}\xi$ и $\operatorname{Im}\xi$ — действительные случайные величины.

Сходимость в с.к. обладает следующими свойствами:

1). $\mathbf{E}\,\underset{k\to\infty}{\text{l.i.m.}}\, \xi_k = \lim\limits_{k\to\infty} \mathbf{E}\xi_k$;

2). $\mathbf{E}|\underset{k\to\infty}{\text{l.i.m.}}\, \xi_k|^2 = \lim\limits_{k\to\infty} \mathbf{E}|\xi_k|^2$;

3). если $\xi = \underset{k\to\infty}{\text{l.i.m.}}\, \xi_k$, $\eta = \underset{k\to\infty}{\text{l.i.m.}}\, \eta_k$, то $\mathbf{E}\xi\overline{\eta} = \lim\limits_{k\to\infty} \mathbf{E}\xi_k\overline{\eta_k}$.

9.5.2. Корреляционная функция случайного процесса. Важную роль в теории второго порядка случайных процессов играет корреляционная функция случайного процесса $K(t,s)$, определяемая для случайного процесса $\xi(t) = \operatorname{Re}\xi(t) + i\operatorname{Im}\xi(t)$ следующим равенством:

$$K(s,t) = \mathbf{E}(\xi(t) - \mathbf{E}\xi(t))(\overline{\xi(s)} - \mathbf{E}\overline{\xi(s)}).$$

Например, для вещественного винеровского процесса при $0 < s < t$ корреляционная функция равна

$$\mathbf{E}\xi(t)\xi(s) = \mathbf{E}[(\xi(t)-\xi(s))+(\xi(s)-\xi(0))(\xi(s)-\xi(0)) = \mathbf{E}(\xi(s)-\xi(0))^2 = s;$$

здесь используется независимость приращений винеровского процесса. В общем виде

$$\mathbf{E}\xi(t)\xi(s) = \min(t,s), \quad t > 0, \ s > 0.$$

Аналогично для пуассоновского процесса с началом в нуле, для которого

$$\mathcal{P}(\xi(t) = k) = \frac{(\lambda t)^k}{k!}e^{-\lambda t}, \quad \mathbf{E}\xi(t) = \lambda t,$$

корреляционная функция $K(t, s)$ также равна

$$K(t, s) = \min(t, s), \quad t > 0, \ s > 0.$$

Далее не нарушая общности будем считать математическое ожидание случайного процесса равным нулю.

9.5.3. Непрерывность в среднем квадратичном. Случайный процесс $\xi(t)$ называется непрерывным в с.к. в точке t, если

$$\xi(t) = \underset{h\to 0}{\text{l.i.m.}}\, \xi(t + h),$$

т.е. (если вспомнить определение сходимости в с.к.) если выполнено $\mathbf{E}|\xi(t + h) - \xi(t)|^2 \to 0$ при $h \to 0$.

Связь между с.к. непрерывностью и корреляционной функцией случайного процесса дается следующими утверждениями:

1. Случайный процесс непрерывен в с.к. в точке t_0 тогда и только тогда, когда его корреляционная функция $K(t, s)$ непрерывна в точке (t_0, t_0).

2. Если $K(t, s)$ непрерывна на диагонали $t = s$, то она непрерывна и всюду.

Заметим, что следствием этих утверждений является с.к. непрерывность процессов Винера и Пуассона, несмотря на то, что реализациями пуассоновского процесса являются кусочно постоянные монотонно неубывающие функции.

9.5.4. Дифференцируемость и интегрируемость в среднем квадратичном. Случайный процесс $\xi(t)$ называется дифференцируемым в с.к. в точке t, если существует с.к. предел $\xi(t + h) - \xi(t)/h$ при $h \to 0$. Этот предел носит название с.к. производной случайного процесса $\xi(t)$:

$$\xi'(t) = \underset{h\to 0}{\text{l.i.m.}}\, \frac{\xi(t + h) - \xi(t)}{h}.$$

Случайный процесс $\xi(t)$ с.к. дифференцируем в точке t тогда и только тогда, когда функция

$$Q(h, k) = \frac{1}{hk}\mathbf{E}(\xi(t + h) - \xi(t))(\overline{\xi(t + k)} - \overline{\xi(t)}) \tag{9.19}$$

сходится при h, $k \to 0$. Отношение (9.19) представляет собой разностное отношение для второй производной. Если предел (9.19) существует и равен второй смешанной производной корреляционной

функции $K(t,s)$ случайного процесса $\xi(t)$, взятой в точке $s = t$. Для существования этого предела достаточно непрерывности второй производной $\dfrac{\partial^2 K(t,s)}{\partial t \partial s}$ в точке $t = s$. В этом случае $\dfrac{\partial^2 K(t,s)}{\partial t \partial s}$ является корреляционной функцией для случайного процесса $\xi'(t)$, являющегося с.к. производной от $\xi(t)$.

Случайный процесс интегрируем в с.к. по Риману на отрезке $[a,b]$, если последовательность римановских интегральных сумм сходится в с.к. при стремлении максимального размера интервала разбиения отрезка $[a,b]$ к нулю.

Для с.к. интегрируемости по Риману случайного процесса достаточно интегрируемости по Риману его корреляционной функции на области $[a,b] \otimes [a,b]$.

Доказательство всех утверждений, связывающих свойства с.к. непрерывности, с.к. дифференцируемости и с.к. интегрируемости случайных процессов со свойствами их корреляционных функций основаны на критериях с.к. сходимости, приведенных в разделе 9.5.2.

Теория второго порядка лежит в основе прогнозирования значения случайного процесса $\xi(t)$ по наблюдению за ним в момент времени $s < t$, рассматриваемому в следующем разделе.

Глава 10

Измерение и прогнозирование

Часто на практике либо правая часть эволюционных уравнений, либо начальные условия оказываются неизвестными, и тем не менее требуется высказаться о состоянии системы в будущем, зная ее поведение, например, на некотором интервале времени $[t_1, t_2]$. Информацию о состоянии системы мы получаем на основании измерения состояния системы, а значит, эта информация содержит некоторую погрешность, определяемую точностью измерений. В этой главе мы остановимся на математических методах теории измерительно-вычислительных систем. Эти методы разработаны под руководством проф. Ю.П. Пытьева [12], [13]; они дают наиболее точный прогноз состояния системы на основе заданной математической модели и дополняющих ее данных измерительных экспериментов, выполненных с некоторой погрешностью.

10.1. Теория измерительно-вычислительных систем

Наиболее точное описание динамики системы получилось бы, если в результате наблюдения за системой была бы построена точная ее математическая модель, состоящая, например, из точных эволюционных уравнений и начальных данных. Однако абсолютно точные измерения принципиально невозможны, любое измерение обязательно содержит погрешность. Кроме того, при всяком измерении измерительный прибор обязательно должен вступить во взаимодействие с изучаемой динамической системой, и это взаимодействие так или иначе изменяет систему; в результате показания прибора имеют отношение не к состоянию “свободной” системы, невозмущенной измерением, а к

состоянию возмущенной системы, в которую “включен” посторонний объект — измерительный прибор.

Поэтому важно знать, насколько существенным оказывается влияние этих измерительных погрешностей на прогноз состояния динамической системы в будущем, а для этого требуется применение методов прогнозирования, предназначенных для получения наиболее точных оценок состояния системы в заданные моменты времени. Эти методы составляют теорию измерительно-вычислительных систем.

Измерительно-вычислительная система, как и следует из ее названия, состоит из двух компонент — измерительной и вычислительной.

Обычно первая является преобразователем специфического для измерения воздействия, — радиационного, теплового, механического или какого-либо другого, — в электрический сигнал. Далее в вычислительной компоненте электрический сигнал подвергается математическому преобразованию, результат которого и дает максимально точный прогноз состояния системы.

Оказывается, соединение измерительных датчиков с интеллектуальной частью — вычислителем, действующим по программе, основанной на анализе математической модели процесса измерения и модели динамической системы, в ряде случаев позволяет преодолеть те ограничения принципиального характера, которые кладут предел совершенствованию измерительной аппаратуры путем чисто технических решений, без применения математики. На уровне измерительно-вычислительной системы все выглядит принципиально по-другому, поскольку вычислительная компонента обычно позволяет, грубо говоря, вычислять и математически моделировать то, что непосредственно ненаблюдаемо.

Главная цель методов теории измерительно-вычислительных систем применительно к рассматриваемой проблеме состоит в том, чтобы получить наиболее точный прогноз состояния динамической системы в ее естественном виде, т.е. прогноз состояния системы, не содержащей измерительный прибор, взаимодействие с которым могло бы исказить ее состояние. Что для этого нужно?

Во-первых, если требуется компенсировать искажения, вызванные взаимодействием с измерительным прибором, то необходимо хорошо представлять себе, к каким именно искажениям динамической системы приводит это взаимодействие. Для этого в теории измерительно-вычислительных систем строится математическая модель, описывающая поведение динамической системы, измерительного прибора и их взаимодействий.

Во-вторых, поскольку нас интересует не возмущенное измерением состояние динамической системы, то необходима ее математическая модель, описывающая ее поведение, не искаженное измерением, а также дающая связь между возмущенным и невозмущенным состо-

яниями системы.

В-третьих, нужны сами математические методы, которые на основании этих моделей дают оценки параметров невозмущенной системы, причем обладающие максимальной точностью.

Наконец, в четвертых, поскольку результат работы измерительно-вычислительной системы, т.е. наиболее точный прогноз состояния невозмущенной динамической системы, в значительной степени определяется тем, насколько близки к действительности используемые математические модели, необходимы методы проверки адекватности моделей. Для этого служат методы теории надежности, которые не только отвечают на вопрос, насколько модель согласуется с полученными в эксперименте данными, но и на вопрос, можно ли использовать заданные модели для того чтобы получить прогноз с гарантированной точностью.

10.1.1. Несколько неформальных определений. Принципы, на которых основана измерительно-вычислительная система, могут быть пояснены следующим образом. Как уже было отмечено, в процессе измерения динамическая система, в которую внедрен измерительный прибор, оказывается в той или иной степени возмущенной. Поскольку при прогнозе могут интересоваться состоянием системы, невозмущенной измерением, на практике получил распространение подход, согласно которому измерительный прибор должен как можно меньше возмущать систему. Концепция измерительно-вычислительной системы основана на новом принципе измерений: в процессе измерения система может претерпевать существенные возмущения, но “на выходе” измерительно-вычислительной системы должны быть получены максимально точные значения параметров изучаемой динамической системы, невозмущенной измерением.

Отметим, что при таком подходе, понятия “исследуемая (изучаемая) система” и “измеряемая система” не совпадают.

Найденные значения параметров прогнозируемой системы и оценка погрешности в теории измерительно-вычислительных систем называются интерпретацией измерения и ошибкой интерпретации, соответственно.

Теория измерительно-вычислительных систем включает новое понятие качества измерительной компоненты, существенно отличающееся от традиционного. Дело в том, что, как оказалось, для наиболее качественных измерений измерительный прибор как таковой и как измерительная компонента измерительно-вычислительной системы должны обладать значительно различающимися характеристиками.

Теория измерительно-вычислительных систем позволяет решать задачи оптимального синтеза измерительной компоненты, специально предназначенной для работы в составе измерительно-вычислительной системы и обеспечивающей ее наивысшее качество как средства

прогноза.

Для построения измерительно-вычислительной системы необходима математическая модель динамической системы со включенным в нее измерительным прибором и учитывающая их взаимодействие. Эта модель носит название модели измерения и связывает состояние динамической системы с результатом измерения. Кроме этого, необходима математическая модель, связывающая измеряемый сигнал со значениями параметров динамической системы, свойственными ее невозмущенному, естественному состоянию. Эта модель называется моделью интерпретации измерения. Теория измерительно-вычислительных систем позволяет, используя эти модели и выходной сигнал ее измерительной компоненты, "вычислить" с помощью ее вычислительной компоненты наиболее точную версию прогнозируемого состояния динамической системы и оценить сопутствующую погрешность.

Поскольку все используемые модели непременно содержат неточности, обусловленные приближенным описанием реальных процессов, теория измерительно-вычислительных систем как составные части содержит теорию надежности модели измерения и теорию надежности интерпретации измерения. Надежность модели измерения характеризует ее адекватность реальному положению вещей, надежность интерпретации измерения позволяет оценить состоятельность интерпретации измерения, т.е. найденных значений прогнозируемых параметров исследуемой динамической системы и оценки погрешности, и тем самым охарактеризовать возможность использования обеих моделей для прогноза с гарантированной точностью.

В терминах теории возможностей надежность модели измерения определяется как возможность ошибочно ее отвергнуть на основе результатов измерения и другой достоверной информации, а надежность интерпретации определяется как возможность ошибочно отвергнуть, как неверный, результат интерпретации, т.е. вычисленные значения параметров исследуемого объекта и соответствующей оценки погрешности.

Надежность позволяет контролировать весь процесс анализа и интерпретации измерения, включая диалог с исследователем и вычисления. В режиме диалога с измерительно-вычислительной системой исследователь может изменять и дополнять модели, основываясь на фактах и догадках, ассоциированных с его научным опытом и промежуточными результатами интерпретации, может привлекать дополнительные измерения, в том числе — тестирующие модель. Вычисляя надежность формируемой в диалоге модели, он контролирует непротиворечивость привлекаемой информации и результатов измерения. А если модели измерения и интерпретации предстоит выбрать

из некоторого класса моделей, исследователь выбирает максимально надежные и использует их, если их надежности определяются как приемлемые.

Исследование адекватности математических моделей измерения и интерпретации и подбор их характеристик с целью максимизации надежности составляют то, что в теории измерительно-вычислительных систем называется анализом измерения.

Прежде, чем перейти к конкретным примерам, объясним в общем случае и достаточно схематично, каким классом измерительных экспериментов мы рассматриваем в этой части пособия и какого рода проблемы нас интересуют.

10.1.2. Схема реального измерения. Пусть динамическая система содержит измерительный прибор, сигнал которого формируется за счет взаимодействия с объектами динамической системы. Нас будет интересовать проявление этого взаимодействия, которое будем описывать так, как будто на вход прибора от объекта и среды поступает сигнал, обозначаемый символом f. Прибор, обозначаемый символом A, устроен следующим образом: поступающий на его вход сигнал f он преобразует в показания прибора — сигнал Af. Этот сигнал и измеряется — оцифровывается и поступает исследователю как результат измерения. Но в процессе измерения и оцифровки обязательно возникают погрешности — ошибки, которые будем описывать как некоторую шумовую добавку, случайный сигнал, прибавляемый к выходному сигналу Af прибора A. Таким образом, результат измерения можно записать схематично в виде

$$\xi = Af + \nu. \tag{10.1}$$

Схему (10.1) будем интерпретировать следующим образом: *сигнал ξ есть искаженный шумом ν результат измерения выходного сигнала Af прибора A, на вход которого подан сигнал f от объекта и среды, возмущенных измерением.*

10.1.3. Схема идеального измерения. Опишем теперь схему идеального измерительного эксперимента, в результате которого мы получили бы значения ровно тех параметров динамической системы, которыми мы интересуемся, причем в нужном нам состоянии — например, не возмущенном измерением. Под идеальным прибором будем понимать такой, на вход которого поступает сигнал f от реального объекта и среды, возмущенных измерением, но на выходе появляются абсолютно точные значения нужных нам параметров. Обозначим такой идеальный прибор символом U, а параметры объекта в невозмущенном состоянии — символом u. Тогда схема идеального измерения запишется в виде

$$u = Uf. \tag{10.2}$$

10.1.4. В чем состоит интерпретация измерения. Наша цель состоит в том, чтобы, имея в распоряжении результат измерения — сигнал ξ из (10.1), и зная математические модели схем измерения (10.1) и (10.2), преобразовать сигнал ξ из соотношения (10.1) к виду, наиболее близкому к сигналу u в (10.2).

Преобразование результата измерения ξ к виду, свойственному измерению с помощью идеального прибора, в теории измерительно-вычислительных систем составляет существо проблемы и н т е р п р е т а ц и и   и з м е р е н и я.

10.1.5. Линейная модель измерения. В этой части мы рассмотрим только линейные измерительно-вычислительные системы, т.е. такие, в которых приборы A и U являются линейными. Под линейными приборами понимают такие приборы, которые обладают следующими свойствами.

1. Если амплитуда сигнала f, поступающего на вход линейного прибора A, увеличится в k раз, то во столько же раз увеличится и амплитуда выходного сигнала.

2. Если на вход линейного прибора A поступает сумма двух сигналов f и g, то на выходе прибора A сигнал равен сумме выходных сигналов Af и Ag

Иными словами, схематически для линейного прибора можно записать соотношение: для любых двух сигналов f и g и для любых чисел α и β справедливо соотношение

$$A(\alpha f + \beta g) = \alpha Af + \beta Ag.$$

10.1.6. Интерпретация измерений с помощью линейных измерительно-вычислительных систем. Для линейных измерительно-вычислительных систем, помимо того, что приборы U и A являются линейными, преобразование сигнала ξ, приводящее его к виду, наиболее близкому к выходному сигналу u идеального прибора U, тоже ищется в классе линейных. Обозначим это преобразование символом R, тогда результат интерпретации $R\xi$ имеет вид

$$R\xi = RAf + R\nu = Uf + (RA - U)f + R\nu. \qquad (10.3)$$

Для любого R результат $R\xi$ такого преобразования можно интерпретировать как искаженный шумом $R\nu$ выходной сигнал прибора RA, на вход которого подан сигнал f от возмущенных объекта и среды (если рассматривать среднюю часть равенств (10.3)), или как выходной сигнал Uf идеального прибора U, искаженный “ложным” сигналом $(RA - U)f$ и шумом $R\nu$ (если рассматривать правую часть равенств (10.3)).

Для того, чтобы обеспечить наиболее точную интерпретацию измерения (10.1), *необходимо так подобрать преобразование* R*, чтобы сделать слагаемые* $(RA-U)f$ *и* $R\nu$ *как можно меньшими по величине.*

Поясним далее, как понимать эту последнюю фразу, т.е. уточним, в каком смысле эти слагаемые должны быть как можно меньше. Ведь ни сигнал f, ни шум ν неизвестны, в лучшем случае на практике можно указать из средние характеристики или диапазон изменений. Но прежде, чем перейти к этим проблемам, приведем пример, в котором расшифровывается понятия схем реального измерения (10.1) и идеального измерения (10.2).

10.1.7. Наблюдения с помощью датчика второго порядка. Рассмотрим измерение, осуществляемое с помощью датчика, представляющего собой механическую колебательную систему, движение которой под действием внешней силы $f(\cdot)$ определяется как решение начальной задачи для дифференциального уравнения второго порядка

$$\mu\ddot{x}(t)+2\alpha\dot{x}(t)+\beta x(t) = f(t), \quad 0 < t \le T, \quad x(0) = 0, \quad \dot{x}(0) = 0. \quad (10.4)$$

Параметры μ, α и β определяют соответственно массу, затухание и упругость подвижной системы датчика, ее начальное смещение $x(0)$ и скорость $\dot{x}(0)$ в (10.4) выбраны равными нулю.

Пусть исследуемая динамическая система полностью описывается заданием функции $f(t)$, $t \in [0, T)$, являющейся внешней силой в уравнении движения маятника. Тогда для определения внешней силы $f(t)$ в каждый момент времени $t \in [0, T]$, или какого-либо проявления ее действия, может измеряться значение $\xi(t) = x(t) + \nu(t)$ смещения $x(t)$ с ошибкой $\nu(t)$ или — соответствующего электрического сигнала. В данном случае речь идет о схеме измерения (10.1), записанной в виде

$$\xi(t) = \frac{1}{\mu\omega}\int\limits_0^t e^{-\gamma(t-\tau)}\sin(\omega(t-\tau))f(\tau)\,d\tau + \nu(t), \quad 0 \le t \le T,$$

в которой первое слагаемое есть решение задачи (10.4), $\gamma = \alpha/\mu$, $\omega^2 = \mu\beta - \alpha^2 > 0$.

Пусть теперь датчик закреплен на поверхности некоторого движущегося объекта, и изучаемая динамическая система в момент времени $t \in [0, T)$ описывается координатой $s(t)$. Этот датчик, описываемый дифференциальным уравнением второго порядка (датчик второго порядка), будем использовать для определения смещения $s(t)$, $0 \le t \le T$, объекта. Если подвижная система датчика упруго связана с объектом и искажающим влиянием датчика на движение объекта можно пренебречь, т.е. отождествить измеряемый и исследуемый объекты, вместо модели (10.4) получим следующую начальную задачу:

$$\mu\ddot{x}(t) + 2\alpha\dot{x}(t) + \beta x(t) = \beta s(t), \quad 0 < t \le T, \quad x(0) = 0, \quad \dot{x}(0) = 0. \quad (10.5)$$

Если корпус датчика при измерении жестко связан с объектом, то обычно регистрируется смещение подвижной системы датчика относительно объекта. В этом случае схема измерения имеет вид равенства

$$\xi(t) = x(t) - s(t) + \nu(t), \quad 0 \le t \le T, \tag{10.6}$$

где $x(t) = \dfrac{\beta}{\mu\omega} \displaystyle\int_0^t e^{-\gamma(t-\tau)} \sin(\omega(t-\tau))s(\tau)\,d\tau$ — решение задачи (10.5). А если выходной электрический сигнал пропорционален скорости подвижной системы, то $\xi(t) = \dot{x}(t) - \dot{s}(t) + \nu(t)$, $0 \le t \le T$.

В тех случаях, когда влиянием датчика на динамику объекта пренебречь нельзя, уравнение (10.5) следует рассматривать совместно с уравнением движения объекта. Например, если последний также является колебательной системой второго порядка, упруго связанной с подвижной системой датчика, жестко закрепленного на объекте, динамика последнего описывается условиями:

$$\begin{gathered} m\ddot{s}(t) + 2a\dot{s}(t) + bs(t) = f(t) + \beta\big(v(t) - s(t)\big), \\ 0 < t \le T, \quad s(0) = 0, \quad \dot{s}(0) = 0. \end{gathered} \tag{10.7}$$

В данном случае m — масса объекта и неподвижной части датчика, параметры a и b определяют взаимодействие объекта со “средой“, а движение объекта определяется действием внешней силы $f(\cdot)$ и силы реакции датчика $\beta\big(x(\cdot) - s(\cdot)\big)$.

Система дифференциальных уравнений (10.5), (10.7) (и соответствующих начальных условий) вместе со схемой измерений определяет математическую модель динамической системы, содержащей в своем составе измерительный прибор. Решение системы (10.5), (10.7) может быть записано в виде двух равенств:

$$v(t) = \int_0^T V(t,\tau)f(\tau)\,d\tau, \quad s(t) = \int_0^T S(t,\tau)f(\tau)\,d\tau, \quad 0 \le t \le T, \tag{10.8}$$

в которых $V(\cdot,\cdot)$ и $S(\cdot,\cdot)$ — известные функции, зависящие от параметров μ, α, β, m, a, b. Схема измерения в данном случае выглядит как модификация (10.6), а именно:

$$\xi(t) = \int_0^T \big(V(t,\tau) - S(t,\tau)\big)f(\tau)\,d\tau + \nu(t), \quad 0 \le t \le T, \tag{10.9}$$

где функция $\xi(\cdot)$ представляет результат измерения смещения датчика относительно объекта $v(\cdot) - s(\cdot)$ с ошибкой $\nu(\cdot)$.

Равенство (10.9) совместно с равенствами (10.8), описывающими динамику системы, определяют модель измерения, учитывающую искажения динамики измеряемого объекта, обусловленные реакцией датчика.

10.2. Методы синтеза ИВС как идеальных приборов

10.2.1. Схема измерений и ее математические модели. Рассмотрим измерение сигнала f, проводимое по схеме

$$\xi = Af + \nu, \tag{10.10}$$

в которой входной сигнал f прибора A моделируется вектором f евклидова пространства $\mathcal{R}_N$, прибор A описывается линейным преобразованием, действующим из $\mathcal{R}_N$ в евклидово пространство R_n, заданным матрицей A размера $n \times N$. Измерение выходного сигнала Af прибора A сопровождается аддитивной погрешностью ν (случайным шумом), рассматриваемом в модели измерения (10.10) как случайный вектор евклидова пространства $\mathcal{R}_n$.

В зависимости от того, что известно относительно математических объектов, моделирующих сигналы и приборы в соотношении (10.10), различают ряд математических моделей, описывающих линейную схему измерений (10.10). Дадим их краткое описание.

Модель измерения $[A, \Sigma]$. Будем говорить, что задана модель $[A, \Sigma]$ схемы измерений (10.10), если известно, что погрешность измерения (10.10) ν — случайный вектор пространства $\mathcal{R}_n$ с нулевым математическим ожиданием $\mathbf{E}\nu = 0$ и невырожденной ковариационной матрицей $\Sigma \in (\mathcal{R}_n \to \mathcal{R}_n)$, и задана матрица $A \in (\mathcal{R}_N \to \mathcal{R}_n)$ линейного прибора (оператора) A. Равенство нулю среднего значения шума означает, что в наших измерениях отсутствует систематическая погрешность.

Подчеркнем, что, пользуясь моделью $[A, \Sigma]$, мы предполагаем, что измеряемый в (10.10) входной сигнал f прибора A априори произволен и может описываться любым вектором евклидова пространства $\mathcal{R}_N$. Иными словами, координаты вектора f могут принимать любые значения, от минус до плюс бесконечности, и нет предпочтений одних значений перед другими.

Приведем пример измерения, описываемого моделью $[A, \Sigma]$. Пусть измеряются параметры равноускоренного движения тела, наблюдаются положения тела в моменты времени $t_1, \ldots, t_n$, причем измерения координаты $x(t)$ в различные моменты времени t производятся независимо одно от другого с ошибкой, обладающей нулевым средним

значением и дисперсией, равной σ^2. Обозначив

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} x_0 \\ V_0 \\ a \end{pmatrix}, \quad A = \begin{pmatrix} 1 & t_1 & t_1^2/2 \\ \dots & \dots & \dots \\ 1 & t_n & t_n^2/2 \end{pmatrix}, \tag{10.11}$$

заметим, что схему измерения можно записать в виде

$$\xi = Af + \nu,$$

так как

$$x(t) = x_0 + Vt + at^2/2.$$

Математическую модель этого измерения запишем в виде $[A, \sigma^2 I]$, где матрица A определена в (10.11).

Модель измерения $[A, f_0, F, \Sigma]$. Учесть априорные знания о входном сигнале f в (10.10) можно, задав среднее значение $(f_0)_i$ каждой координаты f_i вектора $f \in \mathcal{R}_N$, разброс ее значений вокруг этого среднего а также связи между соседними координатами. Чтобы сделать это в рамках аккуратной математической модели, мы будем считать, что в каждом измерительном эксперименте наблюдается реализация случайного вектора f с известным математическим ожиданием $\mathbf{E}f = f_0 \in \mathcal{R}_N$ и заданной ковариационной матрицей $F \in (\mathcal{R}_N \to \mathcal{R}_N)$. Кроме того, так же, как и для модели $[A, \Sigma]$, известны матрица A, являющаяся математической моделью измерительного прибора, и ковариационная матрица Σ случайного вектора ν, причем предполагается, что $\mathbf{E}\nu = 0$.

Итак, будем считать, что если задана модель $[A, f_0, F, \Sigma]$, то, помимо матрицы $A \in (\mathcal{R}_N \to \mathcal{R}_n)$, математического ожидания $\mathbf{E}\nu = 0$ и невырожденной матрицы $\Sigma \in (\mathcal{R}_n \to \mathcal{R}_n)$ шума ν, относительно схемы измерения (10.10) известно, что входной сигнал f прибора A описывается случайным вектором евклидова пространства $\mathcal{R}_N$ с заданным математическим ожиданием $\mathbf{E}f = f_0$ и ковариационной матрицей $F \in (\mathcal{R}_N \to \mathcal{R}_N)$. Как правило, в экспериментальных исследованиях входной сигнал f и шум ν независимы. Мы будем считать это положение верным, если не оговорено противное.

Моделями схемы измерений (10.10), включающими в себя априорную информацию о входном сигнале f прибора A статистического характера, широко пользуются, например, в геофизических исследованиях при дистанционном зондировании атмосферы и поверхности Земли. В частности, при изучении высотного профиля температуры или профилей концентрации тех или иных компонентов атмосферы дистанционными методами имеются представительные данные об изучаемых параметрах, полученные в результате непосредственных измерений концентраций и температуры атмосферы на шарах–зондах,

проводимых в течение длительного периода времени. Эти данные, с учетом районов и сезонов наблюдения, используются для построения оценок математического ожидания и ковариационной матрицы параметра, изучаемого дистанционными методами.

10.2.2. Несмещенный синтез выходного сигнала идеального прибора. Вернемся к задаче интерпретации измерений на измерительно-вычислительных системах. Пусть известны математические модели реального и идеального измерений, в частности, задана модель $[A, \Sigma]$ и матрица U размера $m \times N$, действующая из $\mathcal{R}_N$ в $\mathcal{R}_m$: $U \in (\mathcal{R}_N \to \mathcal{R}_m)$. Матрица A задает модель реального измерительного прибора, в частности, описывает искажения, вносимые прибором в регистрируемый сигнал f — его искажения в измерительных цепях и др. Матрица $U \in (\mathcal{R}_N \to \mathcal{R}_m)$ определяет идеальный измерительный прибор. Задача интерпретации измерения на измерительно-вычислительной системе заключается в получении наиболее точной версии сигнала Uf — т.е. прогнозируемых значений параметров объекта, по данным измерения ξ, проведенного согласно схеме (10.10). Для линейного случая эта задача решается путем линейного преобразования R сигнала ξ:

$$R\xi = RAf + R\nu = Uf + (RA - U)f + R\nu. \qquad (10.12)$$

Вектор $R\xi$ будем называть результатом интерпретации измерения ξ, а отличие $R\xi$ от выходного сигнала идеального прибора Uf — погрешностью интерпретации. Нас интересует наиболее точная интерпретация измерения ξ, т.е. такая, которая сопровождается наименьшей погрешностью. Результат $R\xi$, преобразования R, согласно (10.12), можно представить как искаженный шумом $R\nu$ выходной сигнал прибора RA, на вход которого подан сигнал f; он отличается от нужного нам сигнала Uf двумя слагаемыми — “ложным сигналом” $(RA - U)f$ и шумом $R\nu$. Будем выбирать R в (10.12) так, чтобы слагаемые $(RA - U)f$ и $R\nu$, отличающие сигнал $R\xi$ от Uf, были как можно меньше.

Для того, чтобы добиться наиболее точной интерпретации, оценим слагаемые $(RA - U)f$ и $R\nu$ на основании модели $[A, \Sigma]$.

Начнем с оценки ложного сигнала. Если f априори произвольный вектор из $\mathcal{R}_N$, то при выполнении неравенства $RA \neq U$ найдется такой входной сигнал $f_0 \in \mathcal{R}_N$, что слагаемое $(RA - U)f_0 \neq 0$. Тогда, так как f в (10.10) априори не известен и может быть произвольным, в том числе иметь сколь угодно большую (конечную) длину, то значение величины $\|(RA - U)f\|$ слагаемого $(RA - U)f$ неограничено: в силу линейных свойств матрицы $RA - U$, если длина “входного вектора” увеличится в k раз, во столько же раз увеличится и длина “выходного” вектора. Значит, для того чтобы можно было оценить величину погрешности интерпретации $R\xi - Uf$, складывающейся,

согласно (10.12), из двух слагаемых — $(RA-U)f$ и $R\nu$, необходимо выполнение равенства $(RA-U)f=0$ для всех $f\in\mathcal{R}_N$, что приводит к матричному уравнению

$$RA=U \tag{10.13}$$

относительно матрицы R.

Итак, если соотношение (10.13) не выполнено, оценить погрешность интерпретации не представляется возможным, и такая интерпретация измерения ξ не может быть признана удовлетворительной. Таким образом, (10.13) — необходимое условие для того чтобы на выходе прибора RA можно было бы получить сигнал, близкий к Uf для всех $f\in\mathcal{R}_N$.

Если бы матрица A была квадратной и существовала обратная A^{-1}, уравнение (10.13) имело бы единственное решение $R=UA^{-1}$. В общем же случае уравнение (10.13) может как иметь решения, так и не иметь, причем если решение имеется, то оно может оказаться неединственным.

Рассмотрим теперь слагаемое $R\nu$. Если выполнено соотношение (10.13), то погрешность интерпретации, т.е. отличие сигнала Uf от его оценки $R\xi$, определяется только этим слагаемым. Его величина может быть охарактеризована средней длиной случайного вектора $R\nu$, которая равна $\mathbf{E}||R\nu||=||R\Sigma^{1/2}||_2^2=\mathrm{tr}\ \ R\Sigma R^*$ и при заданной модели $[A,\Sigma]$ может быть вычислена для любой матрицы R.

Охарактеризуем отличие измерительно-вычислительной системы от идеального прибора U величиной максимального значения среднего квадрата погрешности интерпретации (далее ее будем называть среднеквадратичной (с.к.) погрешностью интерпретации):

$$h(R,U)=\sup\left\{\mathbf{E}||R\xi-Uf||^2|\,\xi=Af+\nu,\ f\in\mathcal{R}_N\right\}. \tag{10.14}$$

Чтобы получить как можно более точную версию сигнала Uf путем преобразования (10.12) результата измерения ξ, матрицу R определим так, чтобы величина с.к. погрешности интерпретации $h(R,U)$ была как можно меньше, т.е. из условия

$$h(R,U)\sim\min_{R:\,RA=U}, \tag{10.15}$$

согласно которому на измерительно-вычислительной системе будет синтезирован выходной сигнал Uf идеального прибора U с минимальной с.к. погрешностью.

Заметим, что если матрица R_* — решение задачи (10.15), то оценка $R_*\xi$ несмещенно оценивает сигнал Uf, так как математическое ожидание $R_*\xi$ равно оцениваемой величине Uf:

$$\mathbf{E}R_*\xi=Uf;$$

поэтому задача (10.15) называется задачей несмещенного синтеза сигнала Uf.

Решение задачи (10.15) дается следующей теоремой.

Теорема. *При невырожденной матрице* Σ *и любом операторе* $U \in (\mathcal{R}_N \to \mathcal{R}_m)$ *условие*

$$U(I - A^- A) = 0 \tag{10.16}$$

необходимо и достаточно для разрешимости задачи (10.15). *При выполнении* (10.16) *единственным решением этой задачи является матрица*

$$R_* = U\left(\Sigma^{-1/2}A\right)^- \Sigma^{-1/2}, \tag{10.17}$$

при этом $R_*\xi = Uf + R_*\nu$ *и с.к. погрешность редукции равна*

$$h(U) = \mathbf{E}||R_*\xi - Uf||^2 = ||U\left(\Sigma^{-1/2}A\right)^- ||_2^2 = \operatorname{tr} U\left(A^*\Sigma^{-1}A\right)^- U^*. \tag{10.18}$$

Итак, для решения задачи интерпретации измерения требуется весьма ограниченная информация о модели измерений, а именно, достаточно знать матрицы A и Σ, задающие модель $[A, \Sigma]$ схемы измерения, и матрицу U, связывающую исследуемые параметры динамической системы — выходной сигнал идеального прибора U — с входным сигналом f реального измерительного прибора A. В такой неопределенной ситуации, когда входной сигнал f априори может быть любым элементом $\mathcal{R}_N$, результат линейной интерпретации (т.е. оценка $R\xi$ сигнала Uf и точность $h(U)$ оценивания) определяется только этими тремя матрицами, и случайный элемент $R_*\xi$ как решение минимаксной задачи (10.15) будет самой точной в с.к. версией выходного сигнала Uf идеального прибора U (в классе линейных решений).

10.2.3. Синтез идеального прибора с ограничением на энергию шума на его выходе. Заметим, что рассмотренная в разделе 10.2.2 задача (10.15) является задачей наилучшего в среднем квадратичном (с.к.) приближения выходного сигнала Uf прибора U сигналом $R\xi$, причем для того чтобы с.к. погрешность была конечна, необходимо выполнение равенства (10.13). Часто это требование не может быть выполнено, либо оказывается слишком обременительным, поскольку полученная оценка $R\xi$ сигнала Uf будет сопровождаться неприемлемо большой погрешностью $h(U)$ в (10.18). В этом разделе рассматривается задача интерпретации измерений с несколько иной, “приборной” точки зрения, как задача наиболее точного в среднем квадратичном синтеза идеального прибора при ограничении на уровень шума.

Постановка и решение задачи синтеза идеального прибора. Пусть задана математическая модель $[A, \Sigma]$ схемы измерения (10.10) и R — линейное преобразование этого измерения. Его результат $R\xi$, равный

$$R\xi = RAf + R\nu = Uf + (RA - U)f + R\nu, \tag{10.19}$$

интерпретируется как искаженный шумом $R\nu$ выходной сигнал прибора RA, на вход которого подан сигнал f.

Напомним, что в разделе () целью преобразования R был наиболее точный синтез *сигнала* Uf, и в силу произвольности $f \in \mathcal{R}_N$ для того, чтобы оценить вклад в погрешность синтеза от второго слагаемого в правой части (10.19), нам пришлось потребовать выполнения равенства $RA = U$. Однако если целью является синтез *прибора* $U \in (\mathcal{R}_N \to \mathcal{R}_m)$, то критерий качества такого синтеза может не зависеть от (неизвестного) сигнала f, например, если считать, что прибор RA тем ближе к U, чем меньше матрица RA отличается от матрицы U. Близость матриц будем измерять суммой квадратов разностей всех их матричных элементов, то есть величиной $G(R, U) = \operatorname{tr}(RA - U)(RA - U)^*$.

Такой подход к задаче интерпретации не позволяет оценить, насколько сигнал $R\xi$ отличается от Uf, однако можно утверждать, что сигнал $R\xi$ сформирован с помощью прибора RA, отличающегося от заданного U на величину $G(R, U) = \operatorname{tr}(RA - U)(RA - U)^*$, при этом уровень шума $R\nu$ на его выходе оценивается средней энергией $\mathbf{E}\|R\nu\|^2 = \operatorname{tr} R\Sigma R^*$. Заметим, что в приборостроении в паспортных данных так же обычно указываются параметры, отличающие данный *прибор* от идеального (например, неравномерность амплитудно - частотной характеристики, коэффициент нелинейности фазовой характеристики и пр.), а не оценку погрешности *сигнала* на его выходе от сигнала идеального прибора.

Если согласиться с такой интерпретацией сигнала $R\xi$, то следует отметить, что результат интерпретации теперь характеризуется двумя параметрами, имеющими различный смысл: это отличие синтезированного прибора RA от U, описываемое значением величины $G(R, U) = \operatorname{tr}(RA - U)(RA - U)^*$, и средняя энергия шума $R\nu$ на выходе прибора RA, равная $\mathbf{E}\|R\nu\|^2 = \operatorname{tr} R\Sigma R^*$. С точки зрения потребителя более понятной является величина средней энергии шума на входе прибора RA, для нее можно указать естественные ограничения, выбранные из практических соображений: ожидаемый сигнал не должен "утонуть" в шумах, а для этого их энергия не должна превосходить некоторого значения. Оценить влияние отличия прибора RA от U на результат интерпретации сложнее, так как оно зависит от измеряемых сигналов, характеризуется в терминах матричных элементов и может проявляться весьма разнообразно. Поэтому естественным является такая постановка задачи синтеза

прибора RA, при котором при заданном ограничении на уровень шума на его выходе минимизируется отличие RA от U.

Найдем матрицу преобразования $R \in (\mathcal{R} \to \mathcal{U})$ из условия минимума $G(R,U)$ при заданном ограничении на уровень шума $R\nu$, т.е. из вариационной задачи

$$\inf\left\{G(R,U)|R \in (\mathcal{R} \to \mathcal{U}),\ \mathbf{E}||R\nu||^2 \le \epsilon\right\}, \quad \epsilon \ge 0. \tag{10.20}$$

Если R_ϵ — решение (10.20), то $R\xi$ интерпретируется как искаженный шумом $R_\epsilon\nu$ выходной сигнал прибора $R_\epsilon A$, ближайшего к U при заданном ограничении на уровень шума $\mathbf{E}||R_\epsilon\nu||^2 \le \epsilon$. Решение задачи (10.20) дается следующей теоремой:

Теорема. *Если ковариационная матрица Σ шума ν невырождена, то для любого прибора, заданного матрицей U, задача* (10.20) *разрешима при $0 \le \epsilon < \infty$ и ее решение имеет вид*

$$R_\epsilon = \begin{cases} R(\omega) = UA^*S_\omega^{-1}, & \omega = \omega_\epsilon, \quad 0 < \epsilon < \epsilon_0, \\ U\left(\Sigma^{-1/2}A\right)^{-}\Sigma^{-1/2}, & \epsilon \ge \epsilon_0, \\ 0, & \epsilon = 0, \end{cases} \tag{10.21}$$

где $S_\omega = A^*A + \omega\Sigma$, $\epsilon_0 = \operatorname{tr} U\left(A^*\Sigma^{-1}A\right)^{-}U^*$, и $\omega_\epsilon = \omega_\epsilon(A,\Sigma,U)$ — корень уравнения

$$\operatorname{tr} UA^*S_\omega^{-1}\Sigma S_\omega^{-1}AU^* = \epsilon. \tag{10.22}$$

При этом уровень шума $h(R_\epsilon,U) = \mathbf{E}||R_\epsilon\nu||^2$ и отличие

$$g(R_\epsilon,U) = \operatorname{tr}(R_\epsilon A - U)(R_\epsilon A - U)^*$$

прибора $R_\epsilon A$ от U даются равенствами

$$h(R_\epsilon,U) = \begin{cases} \epsilon, & 0 \le \epsilon < \epsilon_0, \\ \epsilon_0, & \epsilon \ge \epsilon_0; \end{cases} \tag{10.23}$$

$$g(R_\epsilon,U) = \begin{cases} \operatorname{tr} U\left(I - A^*S_\omega^{-1}A\right)^2U^*, & \omega = \omega_\epsilon, \quad 0 < \epsilon < \epsilon_0, \\ \operatorname{tr} U\left(I - A^{-}A\right)U^*, & \epsilon \ge \epsilon_0, \\ \operatorname{tr} UU^*, & \epsilon = 0. \end{cases} \tag{10.24}$$

Для того, чтобы воспользоваться этими результатами на практике, нужно, казалось бы, задаться уровнем шума ϵ, решить нелинейное уравнение (10.22) относительно ω и воспользоваться формулами (10.21), (10.23), (10.24). Трудности такого пути ясны — мало того, что непросто получить численное решение уравнения (10.22), может оказаться, что при выбранном ограничении на уровень шума приборная невязка $g(R_\epsilon,U)$ может оказаться слишком велика — при этом результат синтеза — прибор RA — будет слишком далек от

идеала U, или, наоборот, слишком мала, тогда ограничение на уровень шума можно сделать жестче, уменьшив ϵ и за счет этого увеличив $g(R_\epsilon, U)$, и тогда решать уравнение (10.22) придется заново.

На практике используется подход, связанный с оперативной характеристикой задачи (10.20).

Оперативная характеристика задачи синтеза идеального прибора. Качество синтеза прибора U на измерительно-вычислительной системе, основанного на решении вариационной задачи (10.20), задается соотношениями (10.23), (10.24), которые в параметрическом виде определяют зависимость величины "приборной невязки" $q(R_\epsilon, U)$ от уровня шума $h(R_\epsilon, U)$ на выходе измерительно-вычислительной системы. Эта зависимость носит название оперативной характеристики задачи (10.20). Ее параметром выступает ω, $0 \le \omega < \infty$, и задается она соотношениями

$$\begin{aligned} h(R,U) &= h(R(\omega),U), \\ g(R,U) &= g(R(\omega),U), \quad 0 \le \omega < \infty. \end{aligned} \tag{10.25}$$

График оперативной характеристики обычно изображается на плоскости (h, g): задавая значение ω, $0 \le \omega < \infty$, мы с помощью соотношений (10.25) определяем абсциссу $h(R(\omega), U)$ и ординату $g(R(\omega), U)$ точки графика оперативной характеристики. Таким образом, с каждой точкой графика оперативной характеристики однозначно связаны три значения — параметра ω, используемого для вычисления матрицы $R(\omega)$ по формуле (10.21), уровня шумов $h(R, U)$ на выходе прибора RA и значения приборной невязки $g(R, U)$. Выбрав компромиссные значения $h(R, U)$ и $g(R, U)$ путем указания точки на графике $g = g(h)$, мы тем самым задаем и параметр ω, используемый для вычисления матрицы $R(\omega)$ по формуле (10.21).

10.2.4. Синтез выходного сигнала идеального прибора. Априорные данные. В этом разделе мы остановимся на методах синтеза выходного сигнала Uf прибора U по результатам измерения ξ в схеме (10.10), описываемой моделью $[A, f_0, F, \Sigma]$, рассмотренной в разделе 10.2.1. Напомним, что в этом случае известна априорная информация о входном сигнале f измерительного прибора A, заданная в стохастических терминах, а именно, известно, что f — случайный вектор из $\mathcal{R}_N$ с заданными математическим ожиданием $\mathbf{E}f = f_0$ и ковариационной матрицей $F \in (\mathcal{R}_N \to \mathcal{R}_N)$.

Теперь задача состоит в том, чтобы по наблюдению одного случайного вектора ξ высказаться о значении другого, не доступного прямому наблюдению, случайного вектора Uf. По сути, речь идет о наилучшем в среднем квадратичном (с.к.) приближении значения Uf линейной функцией от ξ.

Наилучшая в среднем квадратичном линейная аппроксимация случайных векторов. Пусть $\xi \in \mathcal{R}_n$ и $u \in \mathcal{R}_m$ — случайные векторы, для

которых заданы их средние значения $\mathbf{E}\xi = \xi_0 \in \mathcal{R}_n$, $\mathbf{E}u = u_0 \in \mathcal{R}_m$, ковариационные матрицы $\Sigma_\xi \in (\mathcal{R}_n \to \mathcal{R}_n)$, $\Sigma_u \in (\mathcal{R}_m \to \mathcal{R}_m)$, а связь между ними определяется значением матрицы их взаимных ковариаций $\Sigma_{u,\xi} \in (\mathcal{R}_n \to \mathcal{R}_m)$. Вектор ξ наблюдается в эксперименте, а u наблюдению не поддается. Построим приближенное значение для u, зависящее от наблюдаемого значения ξ, наиболее близкое в с.к. к $u \in \mathcal{R}_m$.

Будем искать приближение к $u \in \mathcal{R}_m$ в виде линейной функции от ξ:

$$\widehat{u} = R\xi + B,$$

выбирая матрицу $R \in (R_n \to \mathcal{R}_m)$ и вектор $B \in \mathcal{R}_m$ так, чтобы $\widehat{u}$ в с.к. было как можно ближе к u, т.е. из условий минимума функции $\mathbf{E}||u - \widehat{u}||^2$ по параметрам R и B. Задача на минимум в этом случае имеет вид

$$\min\{\mathbf{E}||u - R\xi - B||^2 \mid R \in (R_n \to \mathcal{R}_m),\, B \in \mathcal{R}_m\}.$$

Как известно, наилучшей в с.к. линейной аппроксимацией случайного вектора u по наблюдению случайного вектора ξ является случайный вектор

$$\widehat{u} = u_0 + \Sigma_{u,\xi}\Sigma_\xi^{-1}(\xi - \xi_0), \tag{10.26}$$

при этом с.к. погрешность аппроксимации равна

$$\mathbf{E}||u - \widehat{u}||^2 = \operatorname{tr}\{\Sigma_u - \Sigma_{u,\xi}\Sigma_\xi^{-1}\Sigma_{u,\xi}^*\}. \tag{10.27}$$

Заметим, что матрица

$$\Sigma_{\widehat{u}} = \Sigma_u - \Sigma_{u,\xi}\Sigma_\xi^{-1}\Sigma_{u,\xi}^* \tag{10.28}$$

является матрицей ковариации для случайного вектора $\widehat{u}$.

Наилучший в среднем квадратичном линейный синтез выходного сигнала идеального прибора. Применим теперь результаты раздела 10.2.4 к задаче наилучшего в среднем квадратичном линейного синтеза выходного сигнала идеального прибора.

Пусть задана математическая модель $[A, f_0, F, \Sigma]$ схемы измерения (10.10), известен результат измерения ξ и требуется найти наиболее точную версию сигнала Uf, где f — тот же самый сигнал, который подавался на вход прибора A в измерении (10.10). Как уже упоминалось в разделе 10.2.2, речь идет о наилучшей в с.к. линейной аппроксимации вектора $u = Uf$ по результату наблюдения вектора ξ. Найдем математические ожидания и ковариационные матрицы векторов ξ и Uf, пользуясь заданной моделью $[A, f_0, F, \Sigma]$.

Математические ожидания $\mathbf{E}\xi$ и $\mathbf{E}Uf$ легко вычисляются, если воспользоваться свойством линейности этой операции:

$$\mathbf{E}\xi = \mathbf{E}(Af + \nu) = A\mathbf{E}f + \mathbf{E}\nu = Af_0; \quad \mathbf{E}Uf = U\mathbf{E}f = Uf_0.$$

Для вычисления ковариационных матриц Σ_ξ, Σ_{Uf} и взаимно ковариационной матрицы $\Sigma_{Uf,\xi}$ вспомним, что матрица ковариации случайного вектора равна среднему значению внешнего произведения этого случайного вектора на себя. Поэтому

$$\Sigma_\xi = \mathbf{E}(\xi - \mathbf{E}\xi)(\xi - \mathbf{E}\xi)^* = \mathbf{E}(A(f - f_0) + \nu)(A(f - f_0) + \nu)^* = \\ = A\mathbf{E}(f - f_0)(f - f_0)^* A^* + \mathbf{E}\nu\nu^* = AFA^* + \Sigma,$$

$$\Sigma_{Uf} = \mathbf{E}(Uf - \mathbf{E}Uf)(Uf - \mathbf{E}Uf)^* = \mathbf{E}U(f - f_0)(f - f_0)^* U^* = \\ = U\mathbf{E}(f - f_0)(f - f_0)^* U^* = UFU^*,$$

$$\Sigma_{Uf,\xi} = \mathbf{E}(Uf - \mathbf{E}Uf)(\xi - \mathbf{E}\xi)^* = \mathbf{E}U(f - f_0)(f - f_0)^* A^* = \\ = U\mathbf{E}(f - f_0)(f - f_0)^* A^* = UFA^*.$$

Напомним, что здесь векторы f и ν считаются независимыми.

Итак, если задача наилучшего в с.к. линейного синтеза выходного сигнала идеального прибора Uf поставить как задачу линейной аппроксимации вектора Uf, т.е. как задачу на минимум

$$\min\{\mathbf{E}\|Uf - R\xi - B\|^2 \mid R \in (R_n \to \mathcal{R}_m),\ B \in \mathcal{R}_m\},$$

то ее решением будет матрица $R_* = UFA^*(AFA^* + \Sigma)^{-1}$ и вектор $B_* = Uf_0 - R_*Af_0$, так, что наилучшей в с.к. линейной оценкой $\widehat{Uf}$ выходного сигнала Uf идеального прибора U является

$$\widehat{Uf} = Uf_0 + UFA^*(AFA^* + \Sigma)^{-1}(\xi - Af_0), \tag{10.29}$$

при этом с.к. погрешность синтеза сигнала Uf равна

$$\mathbf{E}\|Uf - \widehat{Uf}\|^2 = \operatorname{tr}\{UFU^* - UFA^*(AFA^* + \Sigma)^{-1}AFU^*\}; \tag{10.30}$$

матрица

$$\Sigma_{\widehat{Uf}} = UFU^* - UFA^*(AFA^* + \Sigma)^{-1}AFU^* \tag{10.31}$$

является матрицей ковариации для случайного вектора $\widehat{Uf}$.

10.3. Надежность модели и надежность интерпретации

В этом разделе обсуждаются проблемы анализа эксперимента с точки зрения:

— согласия математической модели эксперимента с результатом измерения;

— возможности использовать заданную приближенную модель для интерпретации измерения с заданной точностью.

10.3.1. Зачем проверять модель измерения. В предыдущем разделе мы убедились в том, что при интерпретации экспериментальных данных центральное место занимает математическая модель схемы измерения. Если оценка сигнала Uf, построенная на основании этой модели, еще зависит от результата измерения ξ, то величина погрешности этой оценки целиком определяется моделью измерений. Значит, запрограммировав должным образом вычислительную часть измерительно-вычислительной системы, мы обязательно получим некоторый вектор $R\xi$ и оценку погрешности интерпретации для любого ξ, подаваемого на вход вычислителя, не зависимо от того, получен он на приборе A, действующем в соответствии с заданной моделью измерения, или нет (например — на неисправном приборе). Если в первом случае результат интерпретации будет иметь отношение к действительности, то во втором случае — нет.

Чтобы понять важность обсуждаемой проблемы, рассмотрим такую гипотетическую ситуацию. Пусть два исследователя используют одно и то же измерение, проводимое по схеме (10.1), но первый исследователь построил его математическую модель $\mathcal{N}$, а второй — некоторую другую модель $\mathcal{M}$. Далее оба они действуют в соответствии с рекомендациями раздела 10.2.1 и получают результаты, которые могут принципиально отличаться один от другого — ведь модели, на основе которого эти результаты получены, разные! Естественно, исследователи начинают спорить друг с другом, у кого результат лучше. При этом ссылка на точность вычисленной оценки оказывается несостоятельной, так как эта точность теоретическая, вычисленная на основании заданной модели и зависит от этой модели, а не от результата измерения. Разобравшись в ситуации, исследователи должны прийти к необходимости указать, насколько их действия по вычислению результатов интерпретации согласуются с реальностью. Поскольку о реальности можно судить только по результатам эксперимента, то речь идет о характеристике согласия модели измерения, вычислительных алгоритмов и т.п. с результатом наблюдения ξ.

Обычно состоятельность математической модели некоторого процесса или явления проверяют путем сравнения должным образом проведенного натурного и вычислительного эксперимента — в нашем случае для этого можно, например, измерить выходной сигнал прибора A, если на его вход подан известный тестовый сигнал и сравнить результат измерения с вычисленным на основе математической модели: если отличие измерения и вычисления может быть объяснено наличием погрешности измерения (шумом) ν с известной ковариационной матрицей (задающей, в частности, дисперсии каждой

координаты вектора ν), то модель не противоречит экспериментальным данным. Однако при решении задач интерпретации измерений на практике проведение натурного измерения с полностью контролируемыми входными данными, зачастую, невозможно. Тем не менее вывод о состоятельности математической модели измерений в ряде случаев может быть основан на анализе лишь одних выходных данных эксперимента.

Надежность математической модели измерения, проводимого по схеме

$$\xi = Af + \nu, \tag{10.32}$$

является характеристикой модели, оценивающей, насколько возможен данный результат ξ в эксперименте (10.32), описываемом заданной моделью, тем самым надежность оценивает согласие модели с результатом измерения, ее состоятельность. Надежность модели может быть определена как в том случае, когда сигнал f в (10.32) неизвестен (априори любой вектор из $\mathcal{R}_N$ или случайный вектор $\mathcal{R}_N$ с заданными математическим ожиданием и ковариационной матрицей), так и в том случае, когда сигнал (тест) f задан. В последней ситуации речь идет о надежности модели измерения при тестировании, и тогда надежность модели согласуется с концепцией проверки состоятельности, традиционно принятой в математическом моделировании.

Заметим, что отвечая положительно на вопрос о согласии модели с экспериментальными данными, мы признаем, что у нас нет оснований к тому, чтобы отвергнуть модель. Несколько более сложен вопрос о том, можно ли утверждать, что используемая модель позволяет получать оценку для Uf с указанной точностью; тем не менее в ряде случаев ответ на него тоже может быть получен. Для этого служит понятие надежности интерпретации измерения, которая является характеристикой согласия полученного результата интерпретации с данными измерения (10.32).

Математический аппарат, лежащий в основе определения и вычисления надежности, основан на методах проверки статистических гипотез. Здесь мы вновь будем считать известными не только математические ожидания и ковариационные матрицы случайных векторов, но и их распределения.

10.3.2. Надежность модели измерения. Предположим, что в модели $[A, \Sigma]$ сомнительным можно считать лишь оператор A, а про шум ν известно, что его можно моделировать случайным вектором из $\mathcal{R}_n$, контролируемым нормальным распределением, $\nu \sim N(0, \Sigma)$, причем для простоты будем считать, что $\Sigma = \sigma^2 I$, т.е. ν — так называемый белый шум. Предположим также, что f — априори произвольный вектор $\mathcal{R}_N$, причем $N < n$, т.е. размерность f меньше, чем число измерений. На практике обычно эти условия выполнены. В таком случае, если f пробегает все евклидово про-

странство $\mathcal{R}_N$, вектор Af пробегает в $\mathcal{R}_n$ линейное подпространство $\mathcal{R}(A) = \{y \in \mathcal{R}_n : y = Af,\ f \in \mathcal{R}_N\}$, называемое пространством значений A. Его размерность не превосходит N. Таким образом, гипотеза, согласно которой $\xi = Af + \nu$ при некоторых $f \in \mathcal{R}_N$ и $\nu \in \mathcal{R}_n$, $\nu \sim N(0, \sigma^2 I)$, эквивалентна гипотезе о равенстве $\xi = a + \nu$ при некотором $a \in \mathcal{R}(A) \subset \mathcal{R}_n$ и $\nu \in \mathcal{R}_n$, $\nu \sim N(0, \sigma^2 I)$. В качестве альтернативной гипотезы, при выполнении которой модель $[A, \sigma^2 I]$ будет заведомо неверна, примем условие $\xi = a + \nu$, $a \notin \mathcal{R}(A)$. Отметим сразу же, что если гипотеза верна, это не означает, что верна и модель $[A, \sigma^2 I]$, ибо $\mathcal{R}(A)$ не определяет однозначно A.

Дело свелось к задаче проверки статистической гипотезы о параметре нормального распределения проекции ξ_2 случайного вектора ξ на ортогональное дополнение $\mathcal{R}^{\perp}(A) \subset \mathcal{R}_n$ к пространству $\mathcal{R}(A)$: при верной гипотезе математическое ожидание ξ_2 равно нулю, а при верной альтернативе отлична от нуля. Для ее решения используется статистика

$$\alpha(\xi) = P\big(\chi^2_{n_2} \geq \|\xi_2\|^2/\sigma^2\big) = \int\limits_{\|\xi_2\|^2/\sigma^2}^{\infty} p(x)\, dx, \qquad (10.33)$$

названная *надежностью модели* $[A, \Sigma]$. Здесь $\chi^2_{n_2}$ — случайная величина, контролируемая распределением Пирсона хи-квадрат с числом степеней свободы, равным n_2, — размерности пространства $\mathcal{R}^{\perp}(A) \subset \mathcal{R}_n$ [12].

Естественная интерпретация надежности $\alpha(\xi)$ может быть дана в терминах теории возможностей, рассмотренной в следующей главе: $\alpha(\xi)$ — возможность ошибиться, отвергая модель $[A,\ \sigma^2 I]$ на основе результата измерения ξ. Малые значения надежности определенно свидетельствуют против модели $[A,\ \sigma^2 I]$. Дело в том, что статистика $\alpha(\xi)$ в случае верной гипотезы равномерно распределена на $[0, 1]$ (т.е. имеет плотность вероятности, равную единице на $[0, 1]$), а если гипотеза неверна, то плотность распределения $\alpha(\xi)$ не ограничена в нуле, и надежность должна принимать преимущественно малые, близкие к нулю значения.

Итак, решая задачу интерпретации измерения на измерительно-вычислительной системе, исследователь получает, кроме результата прогноза в виде оценки состояния системы $R_*\xi$ и погрешности этой оценки $h(U)$, еще и значение $\alpha(\xi)$ надежности модели $[A,\ \Sigma]$, использованной при вычислении $R_*\xi$ и $h(U)$. Если $\alpha(\xi)$ мало, то как модель, так и найденные значения $R_*\xi$ и $h(U)$ следует охарактеризовать как сомнительные. Обычно модель $[A,\ \Sigma]$ содержит некоторые параметры, в точности значений которых исследователь не уверен. Изменяя их, можно повысить надежность модели и для интерпретации

использовать максимально надежную модель, лучше всего согласующуюся с результатом измерения ξ.

10.3.3. Надежность интерпретации. Как было отмечено в связи с обсуждением состоятельности модели измерения $[A, \Sigma]$, хорошее или плохое ее согласие с результатом измерения ξ, вообще говоря, не свидетельствует ни о возможности, ни, соответственно, о невозможности ее использования для решения задачи интерпретации. В этом разделе мы постараемся разобраться в том, каким образом, исходя из результата измерения ξ, ответить на вопрос о применимости модели $[A, \Sigma]$ для получения интерпретации с гарантированной точностью. Модель интерпретации U будем считать априори не вызывающей сомнения.

Проблема состоит том, что измерение ξ на самом деле может отвечать схеме

$$\xi = \widetilde{A} f + \widetilde{\nu}, \tag{10.34}$$

модель которой $[\widetilde{A}, \widetilde{\Sigma}]$ либо неизвестна, либо известна лишь частично, а исследователь, выбрав модель $[A, \Sigma]$ и определив, в соответствии с теорией измерительно-вычислительных систем, оператор редукции R_*, хочет выяснить, можно ли считать, что $R_*\xi$ оценивает Uf с гарантированной точностью (сигнал f априори неизвестен). Иначе говоря, вопрос состоит в следующем: можно ли, судя по результату измерения ξ и принятой модели измерения $[A, \Sigma]$, считать, что

$$\widetilde{\mathbf{E}} \| R_*(\widetilde{A} f + \widetilde{\nu}) - Uf \|^2 \le \delta, \tag{10.35}$$

где δ — заданная исследователем верхняя граница среднеквадратичной погрешности интерпретации $R_*\xi$ как Uf, $\widetilde{\mathbf{E}}$ — знак математического ожидания, вычисленного в предположении, что верна модель $[\widetilde{A}, \widetilde{\Sigma}]$.

Неравенство (10.35) можно переписать в виде

$$\| R_*(\widetilde{A} - A) f \|^2 + \operatorname{tr} R_*(\widetilde{\Sigma} - \Sigma) R_*^* \le \delta - \operatorname{tr} R_* \Sigma R_*^*, \tag{10.36}$$

где $R_* = U(A^* \Sigma^{-1} A)^- A^* \Sigma^{-1}$.

Обозначим $\mathcal{M}_\delta$ класс моделей $[\widetilde{A}, \widetilde{\Sigma}]$ и сигналов $f \in \mathcal{R}_N$, удовлетворяющих неравенству (10.36). Класс $\mathcal{M}_\delta$ обладает следующим важным свойством: если исследователь знает, что измерение ξ (10.34) сформировано так, что модель $[\widetilde{A}, \widetilde{\Sigma}]$ и сигнал f содержатся в $\mathcal{M}_\delta$, то $R_*\xi$ оценивает Uf с погрешностью, не превосходящей δ. При этом исследователя может не волновать вопрос о состоятельности его модели $[A, \Sigma]$, неважно также, согласно какой конкретно модели $[\widetilde{A}, \widetilde{\Sigma}]$ сформирован сигнал ξ (10.34).

Поскольку в распоряжении исследователя имеется лишь результат измерения ξ, распределение вероятностей которого зависит от $\widetilde{A}$, $\widetilde{\Sigma}$ и f как от параметров, $\xi \sim N(\widetilde{A}f, \widetilde{\Sigma})$, то проблема сводится к задаче проверки статистической гипотезы о параметрах распределения ξ, согласно которой $([\widetilde{A}, \widetilde{\Sigma}], f) \in \mathcal{M}_\delta$.

В работе [12] при некоторых дополнительных предположениях получено решение рассматриваемой задачи и дано выражение для надежности $\beta(\xi)$ гипотезы $([\widetilde{A}, \widetilde{\Sigma}], f) \in \mathcal{M}_\delta$ и тем самым — для надежности интерпретации измерения ξ с гарантированной точностью.

Таким образом, если исследователь, полагаясь на модель $[A, \Sigma]$, получает результат интерпретации $R_*\xi$, то, задавая границу δ погрешности интерпретации $R_*\xi$ как значения Uf параметров исследуемого объекта, он может оценить и состоятельность такой интерпретации, вычислив $\beta_\delta(\xi)$.

Глава 11

Нечеткое моделирование динамических систем

11.1. Возможность и вероятность

Как уже обсуждалось в предыдущих разделах, один из стандартных подходов к моделированию динамической системы состоит в описании ее эволюции путем задания начальных значений координат системы и уравнений, определяющих изменение координат со временем.

Если исследуется поведение сложных нелинейных систем, состоящих из многих взаимодействующих между собой компонент, то, как правило, используют их “огрубленное” описание, так как математическое моделирование поведения всей системы путем задания состояний каждой ее составной части приводит к большим размерностям модели. Вычисление координат, задающих состояние сложных нелинейных систем большого числа переменных принципиально обладает неточностями вследствие приближенности вычислительных алгоритмов, неустойчивости систем и невозможности сколь угодно точного задания начальных данных и эволюционных уравнений, не говоря уже о том, что для такого моделирования вычислительные алгоритмы будут чрезвычайно трудоемкими. Так, например, при исследовании идеального газа как ансамбля частиц используются не описания каждой частицы, а так называемые макроскопические параметры — средняя энергия всех частиц (температура газа), плотность газа как среднее число частиц в единице объема, давление и т.п.; при описании динамики общественных и экономических процессов тоже не интересуются мнением и поведением каждого индивидуума или субъекта экономической деятельности, а используют макроскопические параметры, так или иначе выражающие обобщенное, или

усредненное поведение каждой составной части, причем немаловажным является то, что эти макроскопические параметры являются не абстрактными, а наблюдаемыми величинами.

Для такого "огрубленного" моделирования записываются эволюционные уравнения для макроскопических параметров системы, которые получают либо путем усреднения описания системы на микроуровне, как это делается в статистической физике, либо путем создания эвристических моделей, основанных на наблюдениях за параметрами сложных систем, что чаще встречается в экономике, социологии и других науках, где не разработаны строгие и подробные математические модели поведения каждой составной части системы на микроуровне. При первом подходе описание является приближенным, так как в реальных системах обязательно наблюдаются флуктуации параметров, отличающие их значения от средних; эти флуктуации связываются с невозможностью абсолютно точного определения фазовых координат системы, что приводит к необходимости поиска иных принципов моделирования, чем описание систем в виде детерминированных дифференциальных или иных эволюционных уравнений. В частности, благодаря флуктуациям фазовых координат разрушаются детерминированные фазовые траектории, исчезают особые точки, система как бы "прощупывает" фазовое пространство и блуждает вокруг состояния, предсказываемого решением детерминистского эволюционного уравнения. Второй подход также неточен, поскольку основан на рассуждениях по прецеденту, в то время, как каждая система может обладать некоторыми индивидуальными чертами.

При моделировании сложных систем важно уметь отвечать на вопросы, насколько точным является результат того или иного предсказания, насколько возможны и другие сценарии поведения системы, а также насколько можно доверять тем или иным результатам прогноза, причем желательно, чтобы эти ответы касались каждой отдельно взятой системы, а не ансамбля систем.

Наиболее распространен стохастический подход к описанию неточности задания математической модели явления или системы, согласно которому на множестве параметров задается вероятностное распределение, интерпретируемое следующим образом: вероятность того или иного события равна частоте его появления в (бесконечной) последовательности независимых испытаний. Тем самым, для адекватного применения стохастических принципов к моделированию системы необходимо, чтобы наблюдаемая величина (макроскопический параметр) являлась результатом усреднения независимых случайных величин. Такая схема хорошо работает в равновесной статистической физике, где наблюдаемые макроскопические величины — плотность, температура, давление — являются результатами усреднения функций от "микропеременных" по большому числу частиц. Однако если

речь идет о моделировании уникальных проявлений, об описании неповторяющихся экспериментов, об исходах, которые могут реализоваться один раз, или о принятии решения в данной конкретной обстановке, которая вообще может никогда больше не реализоваться, то описывать такие события в терминах вероятности, связанной с их частотой в серии независимых испытаний, неестественно.

Если эксперимент принципиально не может быть повторен многократно, при стохастическом подходе обычно предлагается интерпретация результатов моделирования путем разыгрывания "мысленного эксперимента", результаты которого, однако, теперь уже не связаны непосредственно с описываемой реальностью.

Неадекватность вероятностных моделей проявляется и при описании мнения эксперта, так как его выражение не является стохастически устойчивым: в разные моменты времени эксперт может принимать различные решения при одних и тех же, казалось бы, неизменных ситуациях.

В этом разделе излагается подход к моделированию физических явлений, при котором неточность задания параметров модели описывается в терминах возможности. Этот подход основан на введенном в работах Ю.П.Пытьева [16] математическом понятии меры возможности, заданной на множестве параметров модели и являющейся аддитивной мерой со значениями в полукольце [0,1], в котором операция сложения определена как максимум, а умножения — как минимум. Конструкции меры возможности применяются в настоящем разделе для моделирования неточности представлений о параметрах динамических систем. Если в стандартном подходе поведение динамической системы описывается фазовой траекторией (при детерминированном моделировании) или случайным процессом (при вероятностном моделировании), то при нечетком описании поведение системы предлагается описывать ансамблем траекторий с заданной на ней мерой возможности. Эта мера задает порядок на множестве траекторий, указывающий, какие траектории более возможны, какие — менее, а какие невозможны вообще. Поведение системы при этом можно определить как нечеткий процесс, описываемый начальным распределением возможности и распределением возможностей перехода системы из одного состояния в другое как функции времени, начального и конечного состояний.

Описания тенденций в ранговой шкале достаточно для выбора оптимального поведения системы, например, минимизирующего возможность риска, поэтому на практике для принятия оптимальных решений ранговое описание системы в терминах возможности адекватно.

Теория возможностей прежде всего является естественным обобщением теории ошибок. Действительно, если в теории ошибок

результат измерения представляется множеством возможных значений измеряемой характеристики объекта, то в теории возможностей допускаются градации возможностей тех или иных значений ошибки. С другой стороны, теорию возможностей, позволяющую формально охарактеризовать градации модальностей типа "возможно" и "необходимо", естественно рассматривать и как модель субъективных суждений, в которой представлены в той или иной степени возможные и более-менее необходимые (достоверные) события, а также и другие атрибуты субъективных суждений, такие например, как "некоторые", "почти все", "приблизительно", "довольно точно", "слегка" и т.д., свойственные и научному языку.

Теоретико-возможностные методы, представленные здесь как альтернативные теоретико-вероятностным, существенно отличаются от последних, прежде всего, тем, что возможность события, в отличие от вероятности, оценивающей частоту его появления в регулярном стохастическом эксперименте, ориентирована скорее на относительную оценку истинности данного события (или его предпочтительности) в сравнении с любым другим событием, причем — в ранговой (порядковой) шкале, в которой могут быть представлены лишь отношения "больше", "меньше" или "равно". Можно сказать, что теория возможностей как таковая и теоретико-возможностные методы моделирования инвариантны относительно любого сохраняющего порядок преобразования шкалы значений возможности. Следовательно, возможность, вообще говоря, не имеет непосредственной (событийной, частотной) интерпретации, свойственной вероятности и связывающей ее с экспериментом. Тем не менее теория возможностей позволяет математически моделировать реальность на основе опытных фактов, знаний, гипотез и суждений исследователей и проверять адекватность построенных моделей.

Предлагаемый подход описания динамических систем в терминах возможности привлекателен, в частности, и тем, что в нем легко учитывается мнение экспертов о возможности того или иного поведения системы, тем самым в математическую модель включается и наблюдатель. Это крайне важно, например, для создания современных средств моделирования сложных социальных и экономических систем на макроуровне, основанных на высказываниях эксперта о возможных ситуациях и состояниях системы, о тенденциях ее поведения, о законах, связывающих те или иные параметры системы. Такие средства позволяют формулировать динамические модели в диалоге с компьютером, основываясь на научном опыте, на данных наблюдения и на выводах модели, исправлять или дополнять модель сообразно новым данным. Обратная связь позволяет исследователю углубить свои знания об объекте исследования, указать наиболее возможные сценарии и разобраться в противоречиях.

11.2. Математические основы теории возможностей

11.2.1. Нечеткие множества и события. Алгебра нечетких множеств. Пусть $(X, \mathcal{A})$ — измеримое пространство, т.е. X — некоторое множество элементов, $\mathcal{A}$ — σ-алгебра подмножеств X. Множество (или событие) $A \in \mathcal{A}$ можно задать с помощью его характеристической функции

$$\chi_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases} \qquad x \in X,$$

т.е. $\chi_A(x)$ равна единице тогда и только тогда, когда $x \in A$. Операции объединения и пересечения приводят к следующим операциям над характеристическими функциями:

$$\chi_{A \cup B}(x) = \max(\chi_A(x), \chi_B(x)), \quad x \in X, \tag{11.1}$$

$$\chi_{A \cap B}(x) = \min(\chi_A(x), \chi_B(x)), \quad x \in X. \tag{11.2}$$

Наряду с "обычными" множествами в теории возможностей рассматриваются так называемые нечеткие множества (события), определяемые характеристическими функциями $\mu_B(\cdot)$, определенными для всех $x \in X$ и принимающими значения на интервале $[0,1]$; равенство $\mu_B(x) = \mu_0$ означает, что элемент $x \in X$ принадлежит нечеткому множеству B с возможностью μ_0. Если $\mu_0 = 0$, то элементу x совершенно невозможно содержаться в B, если же $\mu_0 = 1$, то вполне возможно включение $x \in B$. Таким образом, если множество значений функции $\mu_B(\cdot)$ отлично от $\{0,1\}$, то в X содержатся элементы, о которых невозможно определенно сказать, принадлежат ли они B или нет, это и оправдывает название B как нечеткого множества.

В теории возможностей с содержательной точки зрения функция $\mu_B(\cdot)$ лишь устанавливает порядок на множестве X, указывая, каким элементам X более возможно содержаться в B, а каким — менее, или указывая, что возможность включения некоторых элементов в B одинакова. В этом смысле одно и то же нечеткое множество можно задать с помощью различных характеристических функций, переводящихся одна к другой с помощью сохраняющего порядок монотонного преобразования, переводящего отрезок $[0,1]$ в себя.

Объединение и пересечение нечетких множеств A и B определяется путем задания их характеристических функций соотношениями, аналогичными (11.1), (11.2):

$$\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)), \quad x \in X, \tag{11.3}$$

$$\mu_{A\cap B}(x) = \min(\mu_A(x), \mu_B(x)), \quad x \in X, \tag{11.4}$$

Для четкого множества $A \in \mathcal{A}$ с характеристической функцией $\chi_A(\cdot)$ множество $\overline{A}$ элементов, не содержащихся в A, определяется единственным образом своей характеристической функцией $\chi_{\overline{A}}(x) = 1 - \chi_A(x)$, $x \in X$. Множество $\overline{A}$ называется дополнением A до X. Аналогом дополнения для нечеткого множества B может служить такое множество $\overline{B}$, для которого возможность включения $x \in \overline{B}$ тем больше, чем меньше его возможность включения в B, $x \in X$. Это можно осуществить с помощью операции $\neg(\cdot)$, определенной и принимающей значения на множестве чисел $[0,1]$, обладающей следующими свойствами:

$$\begin{gathered} \neg(\neg(a)) = a, \quad a \le b \Rightarrow \neg(a) \ge \neg(b), \\ \neg(\max(a,b)) = \min(\neg(a), \neg(b)), \\ \neg(\min(a,b)) = \max(\neg(a), \neg(b)), \\ a, b \in [0,1]; \quad \neg(0) = 1; \quad \neg(1) = 0. \end{gathered} \tag{11.5}$$

Дополнением нечеткого множества B с характеристической функцией $\mu_B(\cdot)$ назовем нечеткое множество $\neg B$ с характеристической функцией $\mu_{\neg B}(x) = \neg\mu_B(x)$, $x \in X$. Очевидно, для четкого множества $A \subset \mathcal{A}$ множество $\neg A$ совпадает с $\overline{A}$.

В теории возможностей операция $\neg(\cdot)$ может быть выбрана различными способами, лишь бы были выполнены условия (11.5), благодаря которым сохраняется содержательный смысл дополнения — инверсия порядка на множестве элементов X. В частности, свойствами (11.5) обладает операция $\neg(x) = 1 - x$.

11.2.2. Мера возможности и мера необходимости. Обозначим $\mathcal{L}$ полукольцо $[0,1]$ с операциями сложения, определенной как максимум, и умножения, определенной как минимум, и отношением порядка, совпадающим с естественным:

$$\mathcal{L} = ([0,1], \le, \dotplus, \bullet),$$

здесь и далее для новых бинарных операций используются обозначения $a \dotplus b = \max(a,b)$, $a \bullet b = \min(a,b)$, $a, b, \in [0,1]$. Обозначим $\mathcal{L}(X)$ класс $\mathcal{A}$-измеримых функций $\{f(\cdot)\}$, заданных на X, принимающих значения на $\mathcal{L}$, и содержащих

1. вместе с каждой парой функций f_1 и f_2 их “сумму” и “произведение”, определяемые для каждого $x \in X$ соотношениями

$$f_{1 \dotplus 2}(x) = f_1(x) \dotplus f_2(x) \equiv \max\{f_1(x), f_2(x)\},$$

$$f_{1\bullet 2}(x) = f_1(x) \bullet f_2(x) \equiv \min\{f_1(x), f_2(x)\};$$

Это условие означает, что класс $\mathcal{L}(X)$ вместе с функциями f и g содержит и любые их “линейные комбинации”. В этом смысле $\mathcal{L}(X)$ — “линейный класс” относительно операций $+\!\!\!+$ и $\bullet$.

2. вместе с каждой функцией f ее “отрицание” $\neg f$: $\neg f(x) = \neg(f(x))$;

3. вместе с любой последовательностью функций $\{f_n\} \subset \mathcal{L}(X)$ функции $\overset{\infty}{\underset{k=1}{+}} f_k$ и $\overset{\infty}{\underset{k=1}{\bullet}} f_k$, для каждого $x \in X$ заданные формулами

$$\overset{\infty}{\underset{k=1}{+}} f_k(x) = \sup_{k=1,2,\ldots} f_k(x),$$

$$\overset{\infty}{\underset{k=1}{\bullet}} f_k(x) = \inf_{k=1,2,\ldots} f_k(x).$$

Мерой возможности $p(\cdot)$, заданной на $\mathcal{L}(X)$ и принимающей значения в $\mathcal{L}$, назовем функцию, удовлетворяющую условиям

$$\begin{gathered} p((a \bullet f)(\cdot) + (b \bullet g)(\cdot)) = (a \bullet p(f(\cdot))) + (b \bullet p(g(\cdot))), \\ a, b \in [0,1], \quad f(\cdot), g(\cdot) \in \mathcal{L}(X), \end{gathered} \tag{11.6}$$

$$p\left(\left(\overset{\infty}{\underset{n=1}{+}} f_n\right)(\cdot)\right) = \overset{\infty}{\underset{n=1}{+}} p(f_n(\cdot)), \quad \{f_n(\cdot)\} \subset \mathcal{L}(X). \tag{11.7}$$

Если $f(\cdot) \in \mathcal{L}(X)$ представляет собой характеристическую функцию нечеткого множества (события) B, то значение $p(f(\cdot))$ называется возможностью нечеткого события B и обозначается $P(B)$. Если, в частности,

$$f(x) = \chi_A(x) = \begin{cases} 1, & \text{если } x \in A \subset \mathcal{A}, \\ 0, & \text{если } x \notin A \subset \mathcal{A}, \end{cases}$$

т.е. $f(\cdot)$ — характеристическая функция “четкого” множества A, то $p(f(\cdot)) = P(A)$ называется возможностью (“четкого”) события $A \in \mathcal{A}$. Условие (11.6) определяет свойство аддитивности меры возможностей, а (11.7) — свойство ее счетной аддитивности.

Наряду с мерой $p(\cdot)$, в теории возможностей рассматривается двойственная с ней мера необходимости, характеризующая “невозможность того, что событие не происходит”. Формально она может быть получена следующим образом. Рассмотрим двойственное с $\mathcal{L}$ полукольцо $[0,1]$ с операциями сложения, определенной теперь как минимум, и умножения, определенной как максимум, и отношением порядка, обратным естественному:

$$\widehat{\mathcal{L}} = ([0,1], \widehat{\leq}, \widehat{+}, \widehat{\bullet}),$$

где $a \widehat{+} b = \min(a,b)$, $a \widehat{\bullet} b = \max(a,b)$, $a \widehat{\leq} b$, если $a \geq b$, $a, b, \in [0,1]$.

Класс $\widehat{\mathcal{L}}(X)$ определим аналогично классу $\mathcal{L}(X)$, используя операции $\widehat{+}$ и $\widehat{\bullet}$, при этом заметим, что если $f(\cdot) \in \mathcal{L}(X)$, то $\neg f(\cdot) \in \widehat{\mathcal{L}}(X)$, и наоборот, если $f(\cdot) \in \widehat{\mathcal{L}}(X)$, то $\neg f(\cdot) \in \mathcal{L}(X)$.

Мерой необходимости назовем линейную счетно-аддитивную (относительно операций $\widehat{+}$ и $\widehat{\bullet}$) функцию, заданную на $\widehat{\mathcal{L}}(X)$ и принимающую значения в $\widehat{\mathcal{L}}$.

В частности, такими свойствами обладает функция $n(\cdot)$, определенная соотношениями

$$n(f(\cdot)) = \neg p(\neg f(\cdot)), \quad f(\cdot) \in \widehat{\mathcal{L}}(X).$$

Необходимостью $N(B)$ нечеткого события B назовем значение $n(\mu_B(\cdot))$ меры необходимости на характеристической функции множества B, а необходимостью "четкого" события $A \in \mathcal{A}$ — значение $n(\chi_A(\cdot))$. В этом случае для четкого события справедливо равенство

$$N(A) = \neg P(\overline{A}),$$

означающее, что необходимость события A равно "невозможности того, что A не произойдет". Поскольку $\neg(1) = 0$, а $\neg(0) = 1$, то для четких событий необходимость A однозначно определяется его возможностью; для нечетких же событий значение необходимости может оказаться различным в зависимости от выбора операции $\neg(\cdot)$.

Свойства мер возможности и необходимости изучены в работе [16].

Структуры

$$\mathcal{L} = ([0,1], \leq, +, \bullet) \quad \text{и} \quad \widehat{\mathcal{L}} = ([0,1], \widehat{\leq}, \widehat{+}, \widehat{\bullet})$$

в теории возможностей носят название шкал значений возможностей и необходимостей, а тройка $\{X, \mathcal{L}(X), P\}$ по аналогии с теорией вероятностей называется возможностным пространством.

В работах [16] описана процедура продолжения мер возможности на множество $\overline{\mathcal{L}}(X)$ функций, заданных на X и принимающих значение на $\mathcal{L}$, измеримых относительно σ-алгебры *всех* подмножеств множества X. Далее будем считать, что $\mathcal{L}(X)$ совпадает с $\overline{\mathcal{L}}(X)$.

Меру $P(\cdot)$ на классе нечетких событий с характеристическими функциями $\mu_B(\cdot) \in \mathcal{L}(X)$ можно задать ее распределением $\varphi(\cdot) \in \mathcal{L}(X)$ следующим образом:

$$P(B) = \sup_{x \in X} \{\mu_B(x), \varphi(x)\} \equiv (\mu_B, \varphi).$$

В работе [16] показано, что любую меру возможности можно задать таким образом.

11.3. Принцип относительности в теории возможностей

Выше неоднократно было отмечено, что все выводы теории возможностей должны быть инвариантны по отношению к сохраняющему порядок преобразованию шкалы возможностей $\mathcal{L}$ (и соответствующему ему преобразованию шкалы $\widehat{\mathcal{L}}$). Для этого нужно согласованным образом изменить и классы $\mathcal{L}(X)$ и $\widehat{\mathcal{L}}(X)$, и меры $P(\cdot)$ и $N(\cdot)$. Покажем, как это следует сделать формально.

Пусть Γ — группа преобразований $\gamma(\cdot)$ отрезка $[0,1]$ в себя, задаваемых монотонно возрастающими непрерывными функциями, $\gamma(0) = 0$, $\gamma(1) = 1$. Ясно, что шкала $\mathcal{L}$ инвариантна относительно преобразований $\gamma(\cdot) \in \Gamma$. Более того, множество Γ исчерпывает все преобразования отрезка $[0,1]$ в себя, сохраняющие порядок и удовлетворяющих условиям $\gamma(a \mathbin{\dot{+}} b) = \gamma(a) \mathbin{\dot{+}} \gamma(b)$, $\gamma(a \bullet b) = \gamma(a) \bullet \gamma(b)$, $a, b \in [0,1]$.

Группе Γ соответствуют две группы преобразований, Γ_0 и Γ_*. Первая из них — группа монотонных преобразований класса $\mathcal{L}(X)$, она определяется равенством

$$(\gamma_0 f)(x) = \gamma(f(x)), \quad x \in X, \quad \gamma \in \Gamma, \quad f(\cdot) \in \mathcal{L}(X),$$

а вторая — группа преобразований класса мер $p(\cdot)$:

$$(\gamma_* p)(f(\cdot)) = \gamma(p(f(\cdot))), \quad \gamma \in \Gamma, \quad f(\cdot) \in \mathcal{L}(X), \tag{11.8}$$

при этом имеется взаимно однозначное соответствие между преобразованиями из Γ, Γ_0 и Γ_*, а сами преобразования из Γ, Γ_0 и Γ_* сохраняют порядок на $\mathcal{L}$, $\mathcal{L}(X)$ и на классе мер, соответственно.

Соотношение (11.8) называется условием эквивариантности меры $p(\cdot)$ относительно преобразования $\gamma \in \Gamma$ шкалы $\mathcal{L}$.

Обратимся теперь к характеристикам возможности события $\neg B$, состоящего в том, что событие B не наступает. Как было замечено в предыдущем разделе, это событие зависит от операции $\neg(\cdot)$, определенной на $[0,1]$. На классе $\mathcal{L}(X)$ определим операцию $\neg_0(\cdot)$, согласованную с $\neg(\cdot)$:

$$\neg_0 f(x) = \neg(f(x)), \quad x \in X.$$

Очевидно, что соотношение $f(\cdot) \leq g(\cdot)^1$ эквивалентно неравенству $\neg_0 f(\cdot) \geq \neg_0 g(\cdot)$, и следовательно, легко заметить, что класс

$$\widehat{\mathcal{L}}(X) = \{\neg_0 f, \quad f \in \mathcal{L}(X)\},$$

имеет упорядоченность $\widehat{\leq}$.

Построим группу преобразований $\widehat{\Gamma}$ шкалы $\widehat{\mathcal{L}}$ аналогично группе Γ преобразований шкалы $\mathcal{L}$ и свяжем преобразования $\gamma \in \Gamma$ шкалы $\mathcal{L}$ и $\widehat{\gamma} \in \widehat{\Gamma}$ шкалы $\widehat{\mathcal{L}}$ следующим образом. Пусть $a \in \mathcal{L}$, $\widehat{a} \in \widehat{\mathcal{L}}$, и $\widehat{a} = \neg a$. Если γa — образ a, а $\widehat{\gamma}\widehat{a}$ — образ $\widehat{a}$, естественно считать, что они связаны некоторым соотношением: $\widehat{\gamma}\widehat{a} = \neg_{\widehat{\gamma}\gamma} \gamma a$, операция $\neg_{\widehat{\gamma}\gamma}$, осуществляющая эту связь, преобразует шкалу $\gamma\mathcal{L}$ в шкалу $\widehat{\gamma}\widehat{\mathcal{L}}$. Иными словами, преобразования $\gamma(\cdot) \in \Gamma$ и $\widehat{\gamma}(\cdot) \in \widehat{\Gamma}$ индуцируют преобразование $\neg(\cdot) \to \neg_{\widehat{\gamma}\gamma} = \widehat{\gamma}\neg\gamma^{-1}$, а так как $\neg_{\widehat{\gamma}\gamma}\neg_{\widehat{\gamma}\gamma} = 1$, то должно выполняться и равенство $\widehat{\gamma}\neg\gamma^{-1} = \gamma\neg\widehat{\gamma}^{-1}$, т.е. преобразования $\gamma^{-1}\widehat{\gamma}$ и $\neg$ должны коммутировать:

$$\gamma^{-1}\widehat{\gamma}\neg = \neg\gamma^{-1}\widehat{\gamma}.$$

Условия, связывающие меры $p(\cdot)$ и $n(\cdot)$, и условия, определяющие правило преобразования операции $\neg$, индуцированное преобразованиями $\gamma(\cdot)$ и $\widehat{\gamma}(\cdot)$, определяют соотношение эквивариантности меры необходимости $n(\cdot)$:

$$\widehat{\gamma}_* n(\widehat{\gamma}_0 \widehat{f}(\cdot)) = \widehat{\gamma} n(\widehat{f}(\cdot)), \quad \widehat{f} \in \widehat{\mathcal{L}},$$

по отношению к преобразованию $\widehat{\gamma} \in \widehat{\Gamma}$. Функция $\widehat{\gamma}_* n(\cdot)$ определяет меру для любого преобразования $\widehat{\gamma}_*$.

Итак, пусть $\gamma \in \Gamma$ — преобразование шкалы $\mathcal{L}$, а $\widehat{\gamma} \in \widehat{\Gamma}$ — должным образом согласованное с γ и операцией $\neg$ преобразование шкалы $\widehat{\mathcal{L}}$. Они определяют соответствующие преобразования $f \to \gamma_0 f$, $p(f(\cdot)) \to \gamma_* p(f(\cdot))$, $\widehat{f} \to \widehat{\gamma}_0 f$, $n(\widehat{f}(\cdot)) \to \widehat{\gamma}_* n(\widehat{f}(\cdot))$. Принцип относительности в теории возможностей, определяющий содержательное толкование ее выводов, состоит в том, что любые теоретико-возможностные заключения эквивалентны, если в некоторых шкалах их формулировки совпадают. Содержательное истолкование могут иметь только те заключения, которые не изменяются при изменении шкал.

[1] Это соотношение понимается как неравенство для значений функций $g(\cdot)$ и $f(\cdot)$ на всех аргументах $x \in X$.

11.4. Условная мера возможности

Нечеткие события $\mathcal{A}$ и $\mathcal{B}$ с характеристическими функциями $\mu_{\mathcal{A}} \in \mathcal{L}(X)$ и $\mu_{\mathcal{B}} \in \mathcal{L}(X)$ соответственно называются независимыми, если

$$p(\mu_{\mathcal{A}}(\cdot) \bullet \mu_{\mathcal{B}}(\cdot)) = p(\mu_{\mathcal{A}}(\cdot)) \bullet p(\mu_{\mathcal{B}}(\cdot)),$$

или, что то же самое, если

$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A}) \bullet P(\mathcal{B}).$$

Согласно этому определению в случае независимости $\mathcal{A}$ и $\mathcal{B}$ возможности $P(\mathcal{A} \cap \mathcal{B})$ и $P(\mathcal{A} \cup \mathcal{B})$, полностью определяется возможностями $P(\mathcal{A})$ и $P(\mathcal{B})$.

Условной относительно события $\mathcal{B}$ возможностью нечеткого события $\mathcal{A}$ называется решение $P(\mathcal{A}|\mathcal{B})$ уравнения

$$\min\{P(\mathcal{A}|\mathcal{B}), P(\mathcal{B})\} = P(\mathcal{A} \cap \mathcal{B}). \qquad (11.9)$$

Если решений несколько, каждое из них назовем вариантом условной возможности.

Уравнение (11.9) разрешимо относительно $P(\mathcal{A}|\mathcal{B})$ при любых $P(\mathcal{A} \cap \mathcal{B})$ и $P(\mathcal{B})$, так как $P(\mathcal{B}) \geq P(\mathcal{A} \cap \mathcal{B})$ при всех $\mathcal{A}, \mathcal{B}$ с характеристическими функциями из множества $\mathcal{L}(D)$. Всякое его решение можно задать условиями

$$P(\mathcal{A}|\mathcal{B}) \begin{cases} = P(\mathcal{A} \cap \mathcal{B}), & \text{если} \quad P(\mathcal{B}) > P(\mathcal{A} \cap \mathcal{B}) \\ \geq P(\mathcal{A} \cap \mathcal{B}), & \text{если} \quad P(\mathcal{B}) = P(\mathcal{A} \cap \mathcal{B}) \end{cases}.$$

Среди решений уравнения (11.9) имеются такие, которые можно рассматривать как меру возможности, параметрически зависящей от $\mathcal{B}$. В качестве такой меры можно выбрать вариант условной возможности

$$P(\mathcal{A}|\mathcal{B}) = \begin{cases} P(\mathcal{A} \cap \mathcal{B}), & P(\mathcal{B}) > P(\mathcal{A} \cap \mathcal{B}) \\ 1, & P(\mathcal{B}) = P(\mathcal{A} \cap \mathcal{B}) \end{cases}. \qquad (11.10)$$

11.5. Нечеткие элементы

Рассмотрим еще одну конструкцию, которая формально определяется заданием функции $\mu(\cdot) \in \mathcal{L}(Y)$, где Y — некоторое множество;

эта конструкция носит название нечеткого элемента и, так же, как случайный элемент в теории вероятностей, является фундаментальным понятием теории возможностей.

Пусть $\{X, \mathcal{L}(X), P\}$ — возможностное пространство (аналог вероятностного пространства в теории вероятностей), и мера возможностей задана своим распределением $\psi(\cdot) \in \mathcal{L}(X)$.

Определение. Функцию $\xi(\cdot)$, заданную на X со значениями на множестве Y назовем *нечетким элементом* Y.

Рассмотрим функцию $\mu(\cdot) \in \mathcal{L}(Y)$, определенную соотношением

$$\mu(x) = p_\varphi(\{z \in \mathbf{X} : \ \xi(z) = x\}) = \sup\{\varphi(z) \,|\, z \in \mathbf{X} : \ \xi(z) = x\},$$

где $\varphi(\cdot)$ — распределение возможности на X, задающее меру возможности $p_\varphi(\cdot)$ на $\mathcal{L}(X)$. Значение в точке $x \in Y$ функции $\mu \in \mathcal{L}(Y)$, задающей нечеткий элемент ξ, является возможностью того, что выполняется равенство $\xi = x$, $x \in Y$. Функцию $\mu \in \mathcal{L}(Y)$ будем называть распределением возможности нечеткого элемента ξ. В силу замечания о том, что мера возможности может быть задана и для неизмеримых (по Лебегу) функций, дополнительные требования (типа измеримости) на функцию ξ не накладываются.

Таким образом, нечеткий элемент можно полностью охарактеризовать распределением возможностей его значений $\mu_\xi(\cdot) \in \mathcal{L}(Y)$. Будем говорить, что нечеткий элемент ξ евклидова пространства $\mathcal{R}_n$ задан, если задана пара $\{\mu_\xi(\cdot), Y\}$, где $\mu_\xi(\cdot) \in \mathcal{L}(Y)$ — функция, заданная на $Y \subset \mathcal{R}_n$, принимающая значения на $\mathcal{L}$.

Определение. Нечеткие элементы ξ и η, заданные на X и принимающие значения в Y_1 и Y_2 соответственно, назовем *независимыми*, если $\mu_{\xi,\eta}(x, y) = \min\{\mu_\xi(x), \mu_\eta(y)\}$ для всех $x \in Y_1$, $y \in Y_2$, где $\mu_\xi(\cdot)$, $\mu_\eta(\cdot)$ — распределения возможностей нечетких элементов ξ и η соответственно, а $\mu_{\xi,\eta}(\cdot,\cdot)$ — совместное распределение возможности пары (ξ, η).

Определение. *Условным распределением* $\mu_{\xi|\eta}(\cdot|y)$, $y \in Y_2$, нечеткого элемента $\xi \in Y_1$ при условии, что $\eta = y \in Y_2$, называется любое решение уравнения

$$\min\{\mu_{\xi|\eta}(x|y), \mu_\eta(y)\} = \mu_{\xi,\eta}(x, y), \tag{11.11}$$

где $\mu_\eta(y) = \sup\limits_{x \in Y_1}\{\mu_{\xi,\eta}(x, y)\}, \quad y \in Y_2$ — маргинальное распределение $\eta \in Y_2$.

11.6. Нечеткое моделирование

Рассмотрим пример применения конструкций теории возможностей для анализа и интерпретации измерений, регистрируемых датчиками сигналов, порожденных некоторой физической системой. Пусть измерение проводится по схеме

$$\xi = \xi(f, \nu), \tag{11.12}$$

где f — измеряемый сигнал, ν — погрешность измерения (шум), ξ — результат измерения, и ξ, f, ν — нечеткие элементы множеств X, F и N соответственно, тогда модель измерения задается совместным распределением возможности $\mu_{\xi,f}(\cdot,\cdot)$; значение $\mu_{\xi,f}(x,y)$ есть возможность одновременного появления в эксперименте значений $\xi = x$ и $f = y$, $x \in X$, $y \in Y$.

Интерпретация измерения (11.12) состоит в высказывании о нечетких значениях параметров $u \in U$ изучаемой физической системы, информация о которых содержится в измеряемом сигнале f при известном результате ξ измерительного эксперимента (11.12). Модель интерпретации задана, если известно совместное распределение $\mu_{u,f}(z,y)$ возможности значений сигнала $f = y \in Y$ и параметров $u = z \in U$.

В определенном смысле исчерпывающей информацией о возможных значениях параметра сигнала f при наблюдении ξ является условное распределение $\mu_{u|\xi}(z|x)$ возможности, на его основе можно получить оценки значений u, наилучшие в той или иной степени.

Совместное распределение пары нечетких элементов (ξ, u) можно получить, воспользовавшись формулой “полной возможности”, аналогичной формуле полной вероятности в теории вероятностей:

$$\mu_{\xi,u}(x,z) = \sup_{y\in Y} \min\{\mu_{\xi,f}(x,y), \mu_{u,f}(z,y)\}. \tag{11.13}$$

Вариант условной возможности может быть получен из (11.13) в соответствии с определением (11.11). В частности, одним из вариантов условной возможности $\mu_{u|\xi}(z|x)$, согласно (11.11), является $\mu_{u|\xi}(z|x) = \mu_{\xi,u}(x,z)$. Вариант условной возможности, задающий меру, может быть получен в соответствии с соотношением (11.10).

Рассмотрим функцию возможности потерь $l(x,\widehat{x})$, определенную на $\mathcal{R}_n \otimes \mathcal{R}_n$ и принимающую значения на $[0,1]$ так, что равенство $l(x,\widehat{x}) = l_0$ означает, что возможность понести потери при замене значения x нечеткого элемента ξ на $\widehat{x}$ равна l_0. Аппроксимацией $\widehat{x}$ нечеткого элемента ξ по наблюдению нечеткого элемента η назовем

решение вариационной задачи

$$\widehat{x} = \arg\inf_z \left\{ \sup_x \{\mu_{\xi|\eta}(x|y),\, l(x,z)\} \right\}. \tag{11.14}$$

11.7. Гауссовы нечеткие элементы

11.7.1. Определение гауссова нечеткого элемента. Проследить аналогию между случайным элементом теории вероятностей и нечетким элементом теории возможностей достаточно далеко можно только для определенного класса нечетких элементов, которые будем назвать гауссовыми нечеткими элементами.

Определение. Нечеткий элемент евклидова пространства $\mathcal{R}_n$ назовем *гауссовым*, если его распределение имеет вид

$$\mu_\xi(u) = \varphi\{||\Sigma^{-1/2}(u-u_0)||^2\}, \tag{11.15}$$

где $\varphi(\cdot)$ — монотонно убывающая функция, заданная на положительной полупрямой $\mathcal{R}_+ = \{x \geq 0\}$ числовой оси, $\varphi(0) = 1$, $u_0 \in \mathcal{R}_n$, а Σ — невырожденный самосопряженный положительно определенный оператор, действующий из $\mathcal{R}_n$ в $\mathcal{R}_n$.

Элемент $u_0 \in \mathcal{R}_n$ назовем математическим ожиданием, а оператор Σ — ковариационным оператором нечеткого элемента ξ.

Заметим, что определенный таким образом ковариационный оператор нечеткого элемента определен с точностью до произвольного неотрицательного числового множителя. В самом деле, взяв вместо функции $\varphi(\cdot)$ функцию $\varphi_{a^2}(\cdot) = \varphi\left((\cdot)/a^2\right)$, имеем

$$\varphi\{||\Sigma^{-1/2}(u-u_0)||^2\} = \varphi_{a^2}\{||a\Sigma^{-1/2}(u-u_0)||^2\},$$

так что нечеткому элементу ξ можно приписать и ковариационный оператор, равный Σ/a^2 для любых $a \neq 0$. Чтобы избежать этой неоднозначности, будем считать, что функция $\varphi(\cdot)$ и оператор Σ выбраны так, чтобы спектральная норма оператора Σ равнялась единице: $||\Sigma|| = \sup\limits_{||x||=1} ||\Sigma x|| = 1$.

Приведем свойства математического ожидания и ковариационного оператора.

1. Если $\xi \in \mathcal{R}_n$ — нечеткий элемент с математическим ожиданием u_0 и ковариационным оператором $\Sigma_\xi \in (\mathcal{R}_n \to \mathcal{R}_n)$, то нечеткий элемент $\eta = a\xi \in \mathcal{R}_n$, где $a \in \mathcal{R}_1$, $a \in \mathcal{R}_1$, $a \neq 0$, обладает

математическим ожиданием $au_0 \in \mathcal{R}_n$ и ковариационным оператором $\Sigma \in (\mathcal{R}_n \to \mathcal{R}_n)$.

Действительно, если $\mu_\xi(u) = \varphi(\|\Sigma_\xi^{-1/2}(u-u_0)\|^2)$ — распределение возможностей нечеткого элемента ξ, т.е.

$$P(\xi = u) = \varphi(\|\Sigma_\xi^{-1/2}(u-u_0)\|^2),$$

то тогда

$$P(\eta = z) = P(a\xi = z) = P(\xi = \frac{z}{a}) = \varphi(\|\Sigma_\xi^{-1/2}(\frac{z}{a} - u_0)\|^2) =$$

$$= \varphi(\frac{1}{a}\|\Sigma_\xi^{-1/2}(z - au_0)\|^2) = \varphi_{a^2}(\|\Sigma_\xi^{-1/2}(z - au_0)\|^2),$$

где $\varphi_{a^2}(\cdot)$ — монотонно убывающая функция, и $\|\Sigma_\xi\| = 1$.

2. Результат $\eta = B\xi + b$, $b \in \mathcal{R}_n$, а $B \in (\mathcal{R}_n \to \mathcal{R}_n)$, линейного невырожденного преобразования гауссова нечеткого элемента $\xi \in \mathcal{R}_n$ с математическим ожиданием u_0 и ковариационным оператором Σ_ξ, является гауссовым нечетким элементом с математическим ожиданием, равным $Bu_0 + b$, и ковариационным оператором $B\Sigma_\xi B^*/\|B\Sigma_\xi B^*\|$, действующему из $\mathcal{R}_n$ в $\mathcal{R}_n$.

3. В отличие от случайных элементов, нечеткий элемент, равный сумме двух гауссовых нечетких элементов, вообще говоря, не является гауссовым.

11.7.2. Маргинальное распределение. Введем пространство $\mathcal{R}_n \otimes \mathcal{R}_m$, его элементами являются пары (ξ, η), $\xi \in \mathcal{R}_n$, $\eta \in \mathcal{R}_m$. Рассмотрим нечеткий гауссов элемент $\{\mu_{(\xi,\eta)}((\cdot,\cdot)), \mathcal{R}_n \otimes \mathcal{R}_m\}$ евклидова пространства $\mathcal{R}_n \otimes \mathcal{R}_m$, распределение которого задано в виде

$$\mu_{\xi,\eta}(x,y) = \varphi_0 \left\{ \left(\begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}, \begin{pmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{pmatrix}^{-1} \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \right) \right\}, \quad (11.16)$$

где $\Sigma_x \in (\mathcal{R}_n \to \mathcal{R}_n)$, $\Sigma_y \in (\mathcal{R}_m \to \mathcal{R}_m)$ — самосопряженные положительно определенные линейные операторы, $\Sigma_{xy} = \Sigma_{yx}^* \in (R_m \to \mathcal{R}_n)$, а $\mu_0\{\cdot\}$ — непрерывная монотонно убывающая функция своего числового аргумента, заданная на $\mathcal{R}_+$.

Маргинальное распределение $\mu_\xi(\cdot)$ нечеткого элемента ξ евклидова пространства $\mathcal{R}_n$ определено соотношением

$$\mu_\xi(x) = \sup\{\mu_{(\xi,\eta)}((x,y)) | y \in \mathcal{R}_m\}. \quad (11.17)$$

В рассматриваемом случае функция $\mu_{(\xi,\eta)}((\cdot,\cdot))$ определена соотношением (11.16), в котором $\varphi(\cdot)$ монотонно убывает, значит, для нахождения точной верхней грани в (11.17) требуется, чтобы ее аргумент

принимал минимальное значение по $y \in \mathcal{R}_m$. Запишем аргумент z функции $\varphi(\cdot)$ в (11.17) в виде

$$z = \left(\begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}, \begin{pmatrix} Q & G \\ G^* & T \end{pmatrix} \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \right) =$$

$$= (x - x_0, Q(x - x_0)) + 2(x - x_0, G(y - y_0)) + (y - y_0, T(y - y_0)), \tag{11.18}$$

где

$$Q = \left(\Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}\right)^{-1}, \tag{11.19}$$

$$T = \Sigma_y^{-1} + \Sigma_y^{-1}\Sigma_{yx}\left(\Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}\right)^{-1}\Sigma_{xy}\Sigma_y^{-1}, \tag{11.20}$$

$$G = -\left(\Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}\right)^{-1}\Sigma_{xy}\Sigma_y^{-1}. \tag{11.21}$$

и минимум этой квадратичной формы по y достигается, когда выполнено равенство $y = y_0 - T^{-1}G^*(x - x_0)$. Подставляя это значение элемента y в (11.16), получим, что маргинальное распределение нечеткого элемента ξ также является гауссовым с математическим ожиданием

$$x_{marg} = x_0$$

и ковариационным оператором

$$\Sigma_{x,marg} = \frac{Q - GT^{-1}G^*}{\|Q - GT^{-1}G^*\|} = \frac{\Sigma_x}{\|\Sigma_x\|}.$$

Функция $\mu_\xi(\cdot)$, задающая ее распределение, равна

$$\mu_\xi(x) = \varphi_{\|\Sigma_x\|}\{\|\Sigma_x^{-1/2}(x - x_0)\|\}.$$

11.7.3. Условное распределение нечеткого гауссова элемента. Пусть нечеткий гауссов элемент $\{\mu_{(\xi,\eta)}((\cdot,\cdot)), \mathcal{R}_n \otimes \mathcal{R}_m\}$ евклидова пространства $\mathcal{R}_n \otimes \mathcal{R}_m$ обладает распределением, определенным в (11.16), и требуется найти условное распределение нечеткого элемента ξ при условии, что $\eta = y$.

Вообще говоря, условное распределение $\mu_{\xi|\eta}(\cdot|y)$ нечеткого элемента ξ при условии, что нечеткий элемент $\eta = y$, определено как решение уравнения

$$\mu_{(\xi,\eta)}(x, y) = \min\{\mu_{\xi|\eta}(x|y), \mu_\eta(y)\} \tag{11.22}$$

для всех $x \in \mathcal{R}_n$, где $\mu_\eta(\cdot)$ — маргинальное распределение нечеткого элемента η. Так как, в соответствии с определением предыдущего раздела, $\mu_\eta(y) \geq \mu_{(\xi,\eta)}(x, y)$ для всех $x \in \mathcal{R}_n$, то в качестве варианта условного распределения нечеткого элемента ξ при условии $\eta = y$

можно выбрать распределение, задаваемое функцией $\mu_{(\xi,\eta)}(x,y)$ при фиксированном y. Для гауссова нечеткого элемента (ξ,η) она равна

$$\mu_{\xi,\eta}(x,y) = \varphi\left(\left(\begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix}, \begin{pmatrix} Q & G \\ G^* & T \end{pmatrix}\begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix}\right)\right) =$$

$$= \varphi((x-x_0, Q(x-x_0)) + 2(x-x_0, G(y-y_0)) + (y-y_0, T(y-y_0))) =$$

$$= \varphi(||Q^{1/2}(x-x_{0,y})||^2 + ||\Sigma_y^{-1/2}(y-y_0)||^2), \tag{11.23}$$

где $x_{0,y} = x_0 + \Sigma_{xy}\Sigma_y^{-1}(y-y_0)$, а матрицы Q, G и T определены в (11.19)–(11.21). Однако функция $\mu_{\xi,\eta}(x,y)$ как функция аргумента x при фиксированном $y \neq y_0$ не обращается в единицу ни при каких $x \in \mathcal{R}_n$, и тем самым, не может определять распределение гауссова нечеткого элемента. Но при $x = x_{0,y}$ выполнено равенство $\mu_{\xi,\eta}(x_{0,y},y) = \mu_\eta(y)$, и в силу (11.22) условное распределение $\mu_{\xi|\eta}(x|y)$ в точке $x = x_{0,y}$ может принимать любое значение из интервала $[\mu_{\xi,\eta}(x_{0,y},y), 1]$. Выберем его равным единице. Этот вариант решения (11.22) будем называть условным распределением гауссова нечеткого элемента ξ при условии $\eta = y$.

Условное распределение нечеткого гауссова вектора ξ при условии $\eta = y$ обладает математическим ожиданием $x_{0,y}$ и ковариационным оператором $Q/||Q||$, эти параметры будем называть условным математическим ожиданием и условным ковариационным оператором.

Заметим, что в отличие от гауссовых распределений в теории вероятностей, некоррелированность нечетких гауссовых элементов не означает их независимости.

11.7.4. Аппроксимация нечетких гауссовых элементов. Зададим нечеткий гауссов элемент $\{\mu_{(\xi,\eta)}((\cdot,\cdot)), \mathcal{R}_n \otimes \mathcal{R}_m\}$ евклидова пространства $\mathcal{R}_n \otimes \mathcal{R}_m$ с распределением, определенным в (11.16), и пусть требуется высказаться о значении нечеткого элемента ξ при условии, что $\eta = y$.

Если в формуле (11.14) положить

$$l(x,\widehat{x}) = \begin{cases} 1, & x \neq \widehat{x} \\ 0, & x = \widehat{x} \end{cases},$$

то аппроксимация $\widehat{x}$ ненаблюдаемого нечеткого гауссова элемента ξ по наблюдению $\eta = y$ равна

$$\widehat{x} = x_0 + \Sigma_{xy}\Sigma_y^{-1}(y-y_0) \tag{11.24}$$

и совпадает с условным математическим ожиданием элемента ξ при условии $\eta = y$. В этом смысле решение задачи аппроксимации

нечеткого гауссова вектора формально совпадает с решением задачи наилучшей в среднем квадратичном аппроксимации случайного гауссова вектора. Однако если рассмотреть линейную функцию

$$\eta = A\xi + \nu$$

от нечеткого вектора $\xi \in \mathcal{R}_n$, где $A \in (\mathcal{R}_n \to \mathcal{R}_m)$, а $\nu \in \mathcal{R}_m$ — нечеткий гауссов вектор с распределением $\mu_\nu(z) = \varphi(\|\Sigma_\nu^{-1/2} z\|)$, и ξ и ν независимы, то, в отличие от аппроксимации случайных векторов, аппроксимация нечеткого вектора ξ является нелинейной функцией η. Действительно, совместное распределение нечеткой пары $(\xi, \nu) \in \mathcal{R}_n \otimes \mathcal{R}_m$ в силу их независимости равно

$$\mu_{\xi,\nu}(x, z) = \min\{\varphi(\|\Sigma_\nu^{-1/2} z\|), \varphi(\|\Sigma_\xi^{-1/2}(x - x_0)\|),$$

а поскольку равенства $\eta = y$ и $\xi = x$ одновременно выполняются тогда и только тогда, когда $\nu = y - Ax$, то совместное распределение нечеткой пары $(\xi, \eta) \in \mathcal{R}_n \otimes \mathcal{R}_m$ равно

$$\mu_{\xi,\eta}(x, y) = \min\{\varphi(\|\Sigma_\nu^{-1/2}(y - Ax)\|), \varphi(\|\Sigma_\xi^{-1/2}(x - x_0)\|)$$

и не является гауссовым. Решение задачи аппроксимации в этом случае построено в работе [16].

11.7.5. Собственный базис нечеткого гауссова элемента. Рассмотрим задачу на собственные значения ковариационного оператора Σ_ξ нечеткого гауссова элемента $\xi \in \mathcal{R}_n$:

$$\Sigma_\xi e_k = \lambda_k^2 e_k, \quad k = 1, \ldots, n,$$

и пусть $\{e_k\}$ составляют ортонормированный базис $\mathcal{R}_n$, а собственные числа упорядочены следующим образом: $1 = \lambda_1^2 \geq \lambda_2^2 \geq \ldots \geq \lambda_n^2 > 0$. Тогда распределение нечеткого элемента, записанного в базисе $\{e_k\}$, имеет вид

$$\mu_\xi(u) = \varphi\left(\sum_{k=1}^{n} \lambda_k^{-2}(e_k, u)^2\right).$$

Базис $\{e_k\}$ назовем собственным базисом гауссова нечеткого элемента ξ. Рассмотрим его свойства.

1. Обозначим $P \in (\mathcal{R}_n \to \mathcal{R}_n)$ оператор ортогонального проецирования в $\mathcal{R}_n$ на линейную оболочку $\mathcal{L}(e_{i_1}, e_{i_2}, \ldots, e_{i_k})$ собственных векторов $e_{i_1}, e_{i_2}, \ldots, e_{i_k}$ ковариационного оператора гауссова нечеткого элемента $\xi \in \mathcal{R}_n$ и рассмотрим два нечетких элемента $\eta = P\xi$ и $\zeta = (I - P)\xi$. В силу ортогональности проекторов P и $(I - P)$ взаимно ковариационный оператор $\Sigma_{\eta,\zeta} = 0$, поэтому маргинальные распределения векторов η и ζ обладают математическими ожиданиями Pu_0 и

$(I-P)u_0$ и ковариационными операторами $P\Sigma_\xi P$ и $(I-P)\Sigma_\xi(I-P)$ соответственно.

2. Условное математическое ожидание нечеткого элемента η из пространства $\mathcal{L}(e_{i_1}, e_{i_2}, \ldots, e_{i_k})$ при условии $\zeta = y$ равно Pu_0, а условный ковариационный оператор $\Sigma_{(\eta|\zeta)} = P\Sigma_\xi P$. Иными словами, наблюдение нечеткого элемента ζ не влияет на параметры распределения возможностей ортогонального к нему элемента η.

3. Рассмотрим множество гауссовых нечетких элементов $\{\xi_\varphi\}$ с одинаковыми ковариационными операторами, распределения которых $\mu_\varphi(u) = \varphi(||\Sigma^{-1/2}(u-u_0||^2)$ отличаются только видом функции φ. Чем быстрее убывает функция $\varphi(\cdot)$ с ростом своего числового аргумента, тем более определенно априори можно высказаться о значениях нечеткого элемента ξ_φ (В частности, если $\varphi(0) = 1$, $\varphi(x) = 0$ при $x > 0$, то ξ_φ — обычный четкий элемент пространства $\mathcal{R}_n$, равный своему математическому ожиданию). На основании этого рассуждения введем на множестве нечетких гауссовых элементов операцию сравнения по четкости, задающую на нем частичный порядок.

Рассмотрим для этого гауссовы нечеткие элементы ξ_1 и ξ_2 с одинаковыми ковариационными операторами и математическими ожиданиями $u_{0,1}$ и $u_{0,2}$ и распределениями $\mu_1(u) = \varphi(||\Sigma^{-1/2}(u-u_{0,1})||^2)$ и $\mu_2(u) = \varphi(||\Sigma^{-1/2}(u-u_{0,2})||^2)$ соответственно. Будем говорить, что распределение элемента ξ_1 не менее четко, чем распределение ξ_2, и писать $\xi_1 \overline{\leq} \xi_2$, если

$$\varphi_1(x) \leq \varphi_2(x) \tag{11.25}$$

для всех $x \geq 0$. Будем говорить, что распределение элемента ξ_1 более четко, чем распределение ξ_2, и писать $\xi_1 \overline{<} \xi_2$, если неравенство (11.25) является строгим для всех $x > 0$, при которых $\varphi_1(x) > 0$. Если $\xi_1 \overline{\leq} \xi_2$ и $\xi_2 \overline{\leq} \xi_1$, то $\varphi_1(x) = \varphi_2(x)$ для всех $x \geq 0$ и ξ_1 и ξ_2 совпадают по четкости.

Пусть нечеткий элемент $\xi \in \mathcal{R}_n$ имеет распределение

$$\mu_\xi(u) = \varphi(||\Sigma^{-1/2}(u-u_0)||^2),$$

тогда маргинальное распределение нечеткого элемента $\xi_i = (\xi, e_i) \in \mathcal{R}_1$ равно $\mu_i(x) = \varphi(\lambda_i^{-2}((x-(u_0,e_i))^2) = \varphi_{\lambda_i^{-2}}((x-(u_0,e_i))^2)$, $i = 1, \ldots, n$. Так как $\lambda_1^2 \geq \ldots \geq \lambda_n^2$, то $\xi_n \overline{\leq} \xi_{n-1} \overline{\leq} \ldots \overline{\leq} \xi_1$.

4. Для всех проекций (a, ξ) нечеткого гауссова элемента $\xi \in \mathcal{R}_n$ на вектор $a \in \mathcal{R}_n$, $||a|| = 1$, выполнено соотношение

$$\xi_n \overline{\leq} (\xi, a) \overline{\leq} \xi_1.$$

Это утверждение следует из теоремы Пуанкаре о собственных значениях самосопряженных операторов.

Таким образом, проекция $\xi_1 = (e_1, \xi)$ нечеткого гауссова элемента ξ на $e_1 \in \mathcal{R}_n$ является априори наименее определенной, а проекция $\xi_n = (e_n, \xi)$ на $e_n \in \mathcal{R}_n$ — априори наиболее определенной среди всех проекций (a, ξ), $a \in \mathcal{R}_n$, $||a|| = 1$.

5. Рассмотрим условное распределение (11.23) нечеткого гауссова элемента ξ при условии, что $\eta = y$. Четкость условного распределения, понимаемая так, как это определено в предыдущем пункте, зависит от величины $||\Sigma_y^{-1/2}(y - y_0)||$, т.е. зависит от условия (в этом тоже проявляется отличие свойств гауссовых нечетких элементов от случайных, так как для последних дисперсия условного распределения не зависит от условия). Например, если ξ имеет распределение $\mu_\xi(x) = \varphi(||F^{-1/2}(x - x_0)||^2)$ и произведено измерение проекции $\eta = (\xi, a)$ нечеткого элемента ξ на вектор $a \in \mathcal{R}_n$, $||a|| = 1$, то распределение ортогональной проекции $P_a^\perp \xi = \xi - a(a, \xi) \in \mathcal{R}_{n-1}$ вектора ξ на ортогональное дополнение к вектору a при условии $(\xi, a) = y$ равно

$$\mu_{(P_a^\perp \xi|\eta)}(x|y) = \varphi\left(||\Sigma_{(P_a^\perp \xi|\eta)}^{-1/2}(x - x_{(P_a^\perp \xi|\eta)})||^2 + \frac{(y - y_0)}{||Fa||^2}\right),$$

и при фиксированных a и x максимальное по y значение принимает при $y = y_0$.

Для практики важен вопрос о том, какую проекцию (ξ, a) вектора ξ следует измерять для того, чтобы в наибольшей степени уточнить значение вектора ξ. Более формально, это предложение следует понимать так: какую проекцию вектора ξ следует измерить, чтобы условное распределение проекции вектора ξ на вектор $b \in \mathcal{R}_n$, $||b|| = 1$, максимально нечеткое по $y \in \mathcal{R}_1$ и по $b \in \mathcal{R}_n$, $||b|| = 1$, было бы наиболее четким.

Пусть произведено измерение проекции $\eta = (\xi, a)$ вектора ξ, результатом которого является число y. Нечеткая величина $\zeta = (b, \xi)$ при фиксированных $a \in \mathcal{R}_n$ и $b \in \mathcal{R}_n$ имеет наименьшую четкость при $y = y_0 \equiv (a, u_0)$. Выберем теперь такое $b \in \mathcal{R}_n$, $||b|| = 1$, чтобы нечеткий элемент $\eta = (\xi, b)$ при условии, что $(\xi, a) = y_0$, был бы наиболее нечетким. Согласно утверждениям п.4, это происходит, когда b совпадает с собственным вектором оператора

$$\Sigma_{(\xi|\eta)} = F - \frac{Faa^*F}{(a, Fa)} \in (\mathcal{R}_{n-1} \to \mathcal{R}_{n-1}),$$

отвечающего наибольшему собственному значению (здесь для любого вектора $b \in \mathcal{R}_n$ оператор $bb^*/||b||^2 \in (\mathcal{R}_n \to \mathcal{R}_n)$ — ортогональный проектор на $b \in \mathcal{R}_n$: $bb^*/||b||^2 x = b(x, b)/||b||^2$ для любого $x \in \mathcal{R}_n$). Выберем теперь такое $a \in \mathcal{R}_n$, чтобы четкость этого распределения была наибольшей. Так как четкость гауссова распределения тем выше, чем

меньше это собственное число (совпадающее со спектральной нормой оператора $\Sigma_{(\xi|\eta)} = F - Faa^*F/(a, Fa)$), то задача свелась к поиску такого $a \in \mathcal{R}_n$, $\|a\| = 1$, при котором

$$\left\| F - \frac{Faa^*F}{(a, Fa)} \right\| \sim \inf_{\|a\|=1} .$$

Пусть $Fe_k = \lambda_k^2 e_k$, $1 = \lambda_1^2 \geq \lambda_2^2 \geq \ldots \geq \lambda_n^2$, и $\{e_i\}$ — ортонормированный базис $\mathcal{R}_n$. Если $a = e_1$, то $\|\Sigma_{(\xi|\eta)}\| = \lambda_2^2$; если же $(a, e_1) = 0$, то $\|\Sigma_{(\xi|\eta)}\| = \lambda_1^2$. Проверим, можно ли выбрать a так, чтобы $\|\Sigma_{(\xi|\eta)}\|$ стала меньше, чем λ_2^2. Для этого запишем $\Sigma_{(\xi|\eta)}$ в виде $\Sigma_{(\xi|\eta)} = F^{1/2}(I - \Pi_0)F^{1/2}$, где $\Pi_0 = F^{1/2}aa^*F^{1/2}/(a, Fa)$ — ортогональный проектор на $F^{1/2}a \in \mathcal{R}_n$. Его собственные числа совпадают с собственными числами оператора $(I - \Pi_0)F(I - \Pi_0)$, для которого известно [13], что его наибольшее собственное число лежит на отрезке $[\lambda_1, \lambda_2]$.

Итак, для того, чтобы сделать условное распределение вектора ξ как можно более четким, требуется измерить проекцию ξ на e_1, тем самым уточняя его наименее четкую часть.

11.8. Нечеткая динамика

11.8.1. Нечеткие процессы. Рассмотрим динамическую систему, описываемую набором фазовых координат $\mathbf{q}(t) \in \mathcal{R}_n$, $t \in [0, T]$, для которого задано эволюционное уравнение

$$\dot{\mathbf{q}}(t) = \mathbf{F}(\mathbf{q}, t), \quad t \in (0, T], \tag{11.26}$$

и начальное состояние

$$\mathbf{q}(0) = \mathbf{q}_0.$$

Если $\mathbf{F}(\mathbf{q})$ содержит нечеткие параметры, то естественно считать и его решение $\mathbf{q}(t)$ при фиксированном t нечетким элементом $\mathcal{R}_n$.

По аналогии с теорией случайных процессов, при фиксированном моменте времени t нечеткий процесс будем рассматривать как зависящий от параметра t нечеткий элемент $\{\mathcal{R}, \mu(\cdot, t)\}$, где $\mathcal{R}$ — множество значений координаты $\mathbf{q}(t)$, а $\mu(\mathbf{z}, t)$, $\mathbf{z} \in \mathcal{R}$ — возможность равенства $\mathbf{q}(t) = \mathbf{z}$ в момент времени t; при фиксированном t такая пара задает сечение нечеткого процесса. Однако множество всех сечений еще не определяет нечеткий процесс как функцию времени; для этого следует описать возможность пары равенств $\mathbf{q}(t_1) = \mathbf{z}_1, \mathbf{q}(t_2) = \mathbf{z}_2$ для любых двух моментов времени t_1 и t_2, соответствующей тройки равенств для любых трех моментов времени, и т.д.

Будем говорить, что нечеткий процесс задан, если заданы нечеткие элементы

$$\{\mathcal{R}, \mu(\cdot, t_1)\}, \quad t_1 \in [0, T];$$
$$\{\mathcal{R} \otimes \mathcal{R}, \mu(\cdot, t_1, \cdot, t_2)\}, \quad t_1, t_2 \in [0, T];$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$\{\mathcal{R} \otimes \mathcal{R} \otimes \ldots \otimes \mathcal{R}, \mu(\cdot, t_1, \cdot, t_2, \ldots, \cdot, t_n)\}, \quad t_1, \ldots, t_n \in [0, T];$$

здесь $t_1, t_2, \ldots$ рассматриваются как параметры. Если $\mu(\mathbf{x}, t) = \mu_0$, то возможность того, что координата $\mathbf{q} \in \mathcal{R}_n$ системы в момент времени $t \in [0, T]$ примет значение $\mathbf{x} \in \mathcal{R}$, равна μ_0. Функция $\mu(\mathbf{x}, t)$ рассматривается как распределение возможности нечеткого вектора $\mathbf{q}(t)$, задающего состояние системы в заданный момент времени $t \in [0, T]$.

Пусть $t_1 \leq t_2 \leq \ldots \leq t_n \leq t_{n+1}$. Определим условное распределение $\mu(\cdot, \cdot | \mathbf{x}_1, t_1, \ldots, \mathbf{x}_n, t_n)$ нечеткого элемента $\mathbf{q}(t)$ при условии, что система в моменты времени $t_1, \ldots, t_n$ находилась в состояниях $\mathbf{x}_1, \ldots, \mathbf{x}_n$ соответственно. Оно удовлетворяет уравнению

$$\begin{aligned} \min(\mu(\mathbf{x}_{n+1}, t_{n+1} | \mathbf{x}_1, t_1, \ldots, \mathbf{x}_n, t_n), \mu(\mathbf{x}_1, t_1, \ldots, \mathbf{x}_n, t_n)) = \\ = \mu(\mathbf{x}_1, t_1, \ldots, \mathbf{x}_n, t_n, \mathbf{x}_{n+1}, t_{n+1}). \end{aligned} \tag{11.27}$$

Решение (11.27) определено неоднозначно, однако любое решение (11.27) позволяет по известному распределению возможности $\mu(\mathbf{x}_1, t_1, \ldots, \mathbf{x}_n, t_n)$ вычислить $(n+1)$-точечное распределение возможности $\mu(\mathbf{x}_1, t_1, \ldots, \mathbf{x}_n, t_n, \mathbf{x}_{n+1}, t_{n+1})$. Любое решение (11.27) называется вариантом условного распределения.

11.8.2. Марковские нечеткие процессы. Марковским называется нечеткий процесс, для которого

$$\mu(\mathbf{x}_{n+1}, t_{n+1} | \mathbf{x}_1, t_1, \ldots, \mathbf{x}_n, t_n) = \mu(\mathbf{x}_{n+1}, t_{n+1} | \mathbf{x}_1, t_1).$$

Функцию $\mu(\mathbf{x}_{n+1}, t_{n+1} | \mathbf{x}_n, t_n)$ назовем переходной возможностью.

Для марковских процессов n-точечное распределение возможностей можно получить, зная лишь начальное и переходное распределения возможностей:

$$\begin{aligned} \mu(\mathbf{x}_1, t_1, \ldots, \mathbf{x}_n, t_n) = \\ = \min\{\mu(\mathbf{x}_1, t_1), \mu(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1), \ldots, \mu(\mathbf{x}_n, t_n | \mathbf{x}_{n-1}, t_{n-1})\}. \end{aligned}$$

В соответствии со свойствами условной возможности, запишем соотношение для распределения возможности состояния системы в момент времени t в виде

$$\mu(\mathbf{x}, t) = \sup_{\mathbf{z} \in \mathcal{R}} \big\{ \min\{\mu(\mathbf{x}, t | \mathbf{z}, t_0), \mu(\mathbf{z}, t_0)\} \big\}. \tag{11.28}$$

11.8.3. Уравнение Смолуховского для нечетких процессов. Воспользовавшись формулой полной возможности, получим нечеткий аналог уравнения Смолуховского, известного в теории случайных процессов:

$$\mu(\mathbf{x}, t|\mathbf{x}_0, t_0) = \sup_{\mathbf{z}} \big\{ \min\{\mu(\mathbf{x}, t|\mathbf{z}, \tau), \mu(\mathbf{z}, \tau|\mathbf{x}_0, t_0)\}\big\}, \tag{11.29}$$

причем (11.29) выполнено для любых $\mathbf{x}, \mathbf{x}_0 \in \mathcal{R}_n$, $t > t_0$, $t, t_0 \in [0, T]$ и любого момента времени $\tau \in (t_0, t)$.

Будем далее полагать, что

$$\mu(\mathbf{x}, t_0|\mathbf{x}_0, t_0) = \begin{cases} 0, & \text{если} \quad \mathbf{x} \neq \mathbf{x}_0 \\ 1, & \text{если} \quad \mathbf{x} = \mathbf{x}_0 \end{cases}. \tag{11.30}$$

Укажем класс решений уравнения (11.29), не исчерпывающий, однако, все множество его решений. Рассмотрим функцию $x = f(t; \mathbf{x}_0, t_0)$, определенную для всех $\mathbf{x}_0 \in \mathcal{R}_n$ и для всех t, t_0, таких, что $0 \le t_0 \le t \le T$, принимающую значения в $\mathcal{R}_n$ и такую, что для всех $\mathbf{x}_0 \in \mathcal{R}_n$ и $t_0 \in [0, T]$ выполнено $f(t_0; \mathbf{x}_0, t_0) = \mathbf{x}_0$.

Тогда переходное распределение возможности

$$\mu(\mathbf{x}, t|\overline{\mathbf{x}}, \overline{t}) = \begin{cases} 1, & \text{если} \quad \mathbf{x}(t) = f(t; \mathbf{x}_0, t_0),\ \overline{\mathbf{x}} = f(\overline{t}; \mathbf{x}_0, t_0),\ t_0 \le \overline{t} \le t, \\ 0, & \text{для остальных} \quad \mathbf{x}(t), \overline{\mathbf{x}} \in \mathcal{R}_n,\ t, \overline{t} \in [0, T], \end{cases} \tag{11.31}$$

является решением (11.29). Такие решения будем называть δ-образными, или траекторными, поскольку они, в частности, могут быть получены как решение задачи Коши

$$\begin{aligned} \dot{\mathbf{x}} &= F(\mathbf{x}, t), \\ \mathbf{x}(t_0) &= \mathbf{x}_0, \end{aligned} \tag{11.32}$$

задающей фазовую траекторию динамической системы.

Например, пусть в задаче (11.32) $x \in \mathcal{R}_1$, а правая часть $F(x, t) = \vartheta x$, где ϑ — числовой параметр. Для любого фиксированного ϑ определена траектория

$$\mathbf{x}(t) = f(t; x_0, t_0) = x_0 e^{\vartheta(t - t_0)} \tag{11.33}$$

— δ-образное решение уравнения (11.29) вида (11.31). Если же ϑ рассматривать как нечеткий параметр с распределением $\mu_\vartheta(\cdot)$, то движение по каждой траектории вида (11.33) при $\vartheta = \vartheta_0$ может происходить с возможностью $\mu_\vartheta(\vartheta_0)$, $\vartheta_0 \in \mathcal{R}_1$. В этой ситуации если известно, что в некоторый момент времени τ, $t_0 \le \tau \le t$, система находится в точке с координатой $y = x_0 e^{\vartheta_0(\tau - t_0)}$, $\tau > t_0$, то при $t > \tau$ ее траектория полностью определена.

11.8.4. Однородные нечеткие процессы с независимыми приращениями. Рассмотрим модель нечеткого процесса, в которой нечеткость играет более заметную роль, чем в примере раздела 11.8.3. Эта модель описывает возможность блуждания системы, когда из каждой точки фазового пространства есть возможность движения по нескольким направлениям с различными скоростями.

Заметим, что для построения моделей случайного блуждания в стохастической теории привлекаются соображения как качественного, так и количественного характера, что позволяет построить дифференциальное уравнение для переходной плотности вероятностей, определяющее ее численные значения (при дополнительных условиях — начальных, краевых и т.п.). В теории возможностей же содержательный смысл имеют лишь утверждения "более возможно", "менее возможно", "равновозможно", и, согласно принципу относительности теории возможностей, см. раздел , результат моделирования должен быть инвариантен по отношению к сохраняющим порядок преобразованиям шкалы возможностей.

Рассмотрим систему, состояние которой в любой момент времени $t \geq 0$ описывается нечетким вектором $\mathbf{q}(t) \in \mathcal{R}_n$. Если в момент времени t_0 система находилась в состоянии $\mathbf{x}_0 \in \mathcal{R}_n$, а в момент времени $t_1 > t_0$ — в состоянии $\mathbf{x}_1$, то вектор $\mathbf{x}_1 - \mathbf{x}_0$ назовем приращением нечеткого процесса $\mathbf{q}(t)$, и величину его охарактеризуем числом $\Delta(\mathbf{x}, \mathbf{x}_0) = \|\Sigma^{-1/2}(\mathbf{x} - \mathbf{x}_0)\|$, где $\Sigma \in (\mathcal{R}_n \to \mathcal{R}_n)$ — невырожденный самосопряженный положительно определенный оператор.

Пусть изменение координат системы подчиняется следующим условиям.

а). Возможность изменения координат системы от значения $\mathbf{x}_1$ в момент времени t_1 до $\mathbf{x}_2$ в момент времени t_2 зависит только от величины приращения $\Delta(\mathbf{x}, \mathbf{x}_0)$ и от длины интервала времени $t_2 - t_1$ и не зависит ни от состояния системы в момент времени t_1, ни от положения интервала времени (t_1, t_2) на числовой оси.

б). При фиксированных t_1 и t_2 возможность больших приращений меньше, чем малых, и определяется величиной

$$V(\mathbf{x}_2, t_2, \mathbf{x}_1, t_1) = \frac{\Delta(\mathbf{x}_2, \mathbf{x}_1)}{f(t_2 - t_1)}, \tag{11.34}$$

где $f(\cdot)$ — некоторая функция, определенная и принимающая значения на множестве положительных чисел.

Величина $V(\mathbf{x}_2, t_2, \mathbf{x}_1, t_1)$ имеет смысл "средней скорости" движения системы из состояния $\mathbf{x}_1$ в $\mathbf{x}_2$ за время от t_1 до t_2, если длительность интервала времени принять равным $f(t_2 - t_1)$. Утверждения а)–б) означают, что процесс $\mathbf{q}(t)$ является марковским нечетким

процессом с переходной возможностью, записываемой в виде

$$\mu(\mathbf{x}, t|\mathbf{x}_0, t_0) = \varphi\left(\frac{||\Sigma^{-1/2}(\mathbf{x}-\mathbf{x}_0)||^2}{(f(t-t_0))^2}\right),$$

удовлетворяющей условию (11.30), где $\varphi(\cdot)$ — монотонно убывающая функция числового аргумента, действующая из $\mathcal{R}_+ = \{z \geq 0\}$ в $[0,1]$, такая, что $\varphi(0) = 1$.

Будем считать, что $\varphi(\cdot)$ непрерывная функция, и покажем, что условия а)–б) определяют нечеткий процесс тогда и только тогда, когда $f(\cdot)$ является однородной линейной функцией. Действительно, для переходных возможностей выполнено уравнение Смолуховского (11.29), которое в данном случае при любом фиксированном τ, $t_0 < \tau < t$, имеет вид

$$\varphi\left(\left(\frac{||\Sigma^{-1/2}(\mathbf{x}-\mathbf{x}_0)||}{f(t-t_0)}\right)^2\right) =$$

$$= \sup_{\mathbf{z}\in\mathcal{R}_n} \min\left\{\varphi\left(\left(\frac{||\Sigma^{-1/2}(\mathbf{x}-\mathbf{z})||}{f(t-\tau)}\right)^2\right), \varphi\left(\left(\frac{||\Sigma^{-1/2}(\mathbf{z}-\mathbf{x}_0)||}{f(\tau-t_0)}\right)^2\right)\right\},$$

и в силу непрерывности и монотонности функции $\varphi(\cdot)$ это равенство эквивалентно соотношению

$$\frac{||\Sigma^{-1/2}(\mathbf{x}-\mathbf{x}_0)||^2}{(f(t-t_0))^2} = \inf_{\mathbf{z}\in\mathcal{R}_n} \max\left\{\frac{||\Sigma^{-1/2}(\mathbf{x}-\mathbf{z})||^2}{(f(t-\tau))^2}, \frac{||\Sigma^{-1/2}(\mathbf{z}-\mathbf{x}_0)||^2}{(f(\tau-t_0))^2}\right\}. \tag{11.35}$$

Необходимым, а в данном случае и достаточным условием минимакса в (11.35) является равенство функций аргумента $\mathbf{z} \in \mathcal{R}_n$:

$$\frac{||\Sigma^{-1/2}(\mathbf{x}-\mathbf{z})||^2}{(f(t-\tau))^2} = \frac{||\Sigma^{-1/2}(\mathbf{z}-\mathbf{x}_0)||^2}{(f(\tau-t_0))^2}, \tag{11.36}$$

и противоположная направленность их градиентов. Линиями уровней функций (11.36) являются эллипсоиды с центрами в точках $\mathbf{x}$ и $\mathbf{x}_0$, и точка, в которой достигается минимакс (11.35), является точкой касания этих эллипсоидов. Так как квадратичные формы слева и справа в (11.36) определяются одним и тем же оператором Σ, то точка касания эллипсоидов лежит на прямой, соединяющей их центры. Это приводит к соотношению

$$\mathbf{z} = \alpha\mathbf{x} + (1-\alpha)\mathbf{x}_0, \tag{11.37}$$

где $\alpha \in (0,1)$. Подставив (11.37) в (11.35), с учетом (11.36) найдем

$$\frac{||\Sigma^{-1/2}(\mathbf{x}-\mathbf{x}_0)||^2}{(f(t-t_0))^2} = (1-\alpha)^2\frac{||\Sigma^{-1/2}(\mathbf{x}-\mathbf{x}_0)||^2}{(f(t-\tau))^2} = \alpha\frac{||\Sigma^{-1/2}(\mathbf{x}-\mathbf{x}_0)||^2}{(f(\tau-t_0))^2},$$

что приводит к уравнению для определения функции $f(\cdot)$:

$$\frac{\tau-t_0}{t-t_0} = \frac{f(\tau-t_0)}{f(t-\tau)+f(\tau-t_0)},$$

или, после замены переменных $t' = \tau - t_0$, $t'' = t - \tau$, к соотношению

$$\frac{t'}{t'+t''} = \frac{f(t')}{f(t')+f(t'')},$$

эквивалентному

$$\frac{f(t')}{t'} = \frac{f(t'')}{t''},$$

откуда следует $f(t) = kt$,где $k = \text{const}$, $t \geq 0$.

Итак, условия а)–б) определяют нечеткий процесс $\mathbf{q}(t)$ с переходной возможностью

$$\mu(\mathbf{x},t|\mathbf{x}_0,t_0) = \varphi\left(\frac{||\Sigma^{-1/2}(\mathbf{x}-\mathbf{x}_0)||^2}{(t-t_0)^2}\right). \tag{11.38}$$

Пусть в начальный момент времени $t = 0$ распределение возможностей задано в виде

$$\mu(\mathbf{x}_0,0) = \begin{cases} 1, & \text{при } \mathbf{x}_0 = \mathbf{x}_* \\ 0, & \text{при } \mathbf{x}_0 \neq \mathbf{x}_* \end{cases},$$

а переходная возможность определена в (11.38), тогда

$$\mu(\mathbf{x},t) = \varphi\left(\frac{||\Sigma^{-1/2}(\mathbf{x}-\mathbf{x}_*)||^2}{(t)^2}\right), \tag{11.39}$$

т.е. эволюция системы описывается распространяющейся волной возможности (11.39). Заметим, что при фиксированном t сечение процесса является нечетким гауссовым элементом.

Процессы с переходной возможностью (11.36) будем называть нечеткими процессами с гауссовыми переходами.

Поясним теперь, как можно понимать уравнение (11.26), если считать его правую часть нечетким вектором. Положим $\mathbf{x} = \mathbf{x}_0 + \mathbf{V}(t-t_0)$, где $\mathbf{V} \in \mathcal{R}_n$ — скорость изменения состояния системы

от $\mathbf{x}_0$ в момент времени t_0 до $\mathbf{x}$ в момент времени t. Записав гауссову переходную возможность (11.38) в виде

$$\mu(\mathbf{x}_0 + \mathbf{V}(t - t_0), t|\mathbf{x}_0, t_0) = \varphi(\|\Sigma^{-1/2}\mathbf{V}\|^2),$$

заметим, что для процессов с гауссовыми переходами возможность их эволюции вдоль любого выделенного направления $\mathbf{V}/\|\mathbf{V}\|$ определяется средней скоростью $\mathbf{V}$, которая не зависит от величины приращения $\mathbf{x} - \mathbf{x}_0$. Поэтому вектор $\mathbf{V}$ для этих процессов можно рассматривать как мгновенную скорость изменения состояния — по аналоги с равномерным движением в механике, для которого средняя и мгновенная скорости совпадают. Таким образом, если в (11.26) правая часть является нечетким вектором с распределением $\mu_{\mathbf{F}(\mathbf{q},t)}(\mathbf{V}) = \varphi(\|\Sigma^{-1/2}\mathbf{V}\|^2)$, $\mathbf{V} \in \mathcal{R}_n$, не зависящим от состояния процесса и от времени, то соотношение (11.26) будем рассматривать как указание на то, что задана средняя (и мгновенная) скорость изменения состояния системы, определяющая нечеткий процесс с гауссовыми переходами.

11.8.5. Процессы с дискретным временем и конечным числом состояний. Рассмотрим дискретный марковский нечеткий процесс (нечеткую цепь Маркова), для которого задано распределение возможностей $\mu(x_k, 0) = b_k$, $k = 1, \ldots, N$, начальных состояний и переходная возможность $\mu(x_k, (m+1)\tau|x_j, m\tau) = c_{j,k}^{(m)}$, $k, j = 1, \ldots, N$, $m = 0, 1, 2, \ldots$ Тогда состояние системы в любой момент времени $m\tau$, $m = 1, 2, \ldots$, характеризуется распределением возможностей $\mu(x_k, m\tau) = \mu_k^{(m)}$, удовлетворяющем уравнению Смолуховского

$$\mu_j^{(m)} = \max_i \min\{c_{j,i}^{(m-1)}, \mu_i^{(m-1)}\}, \quad j = 1, 2, \ldots, N, \quad m = 1, 2, \ldots,$$

с начальным условием

$$\mu_i^{(0)} = b_i, \quad i = 1, 2, \ldots, N.$$

Эти соотношения можно рассматривать как линейные относительно операций “$+$” и “$\bullet$”, задающие явную схему расчета:

$$\mu_j^{(m)} = \mathop{+}_{i=1}^{N} c_{j,i}^{(m-1)} \bullet \mu_i^{(m-1)}, \quad j = 1, 2, \ldots, N, \quad m = 1, 2, \ldots,$$

с начальным условием

$$\mu_i^{(0)} = b_i, \quad i = 1, 2, \ldots, N.$$

11.8.6. Волны возможности. Вернемся к модели, в которой состояние системы в любой момент времени $t \in [0, T]$ описывается вектором $\mathbf{x}$ евклидова пространства $\mathcal{R}_n$, и изменение состояния

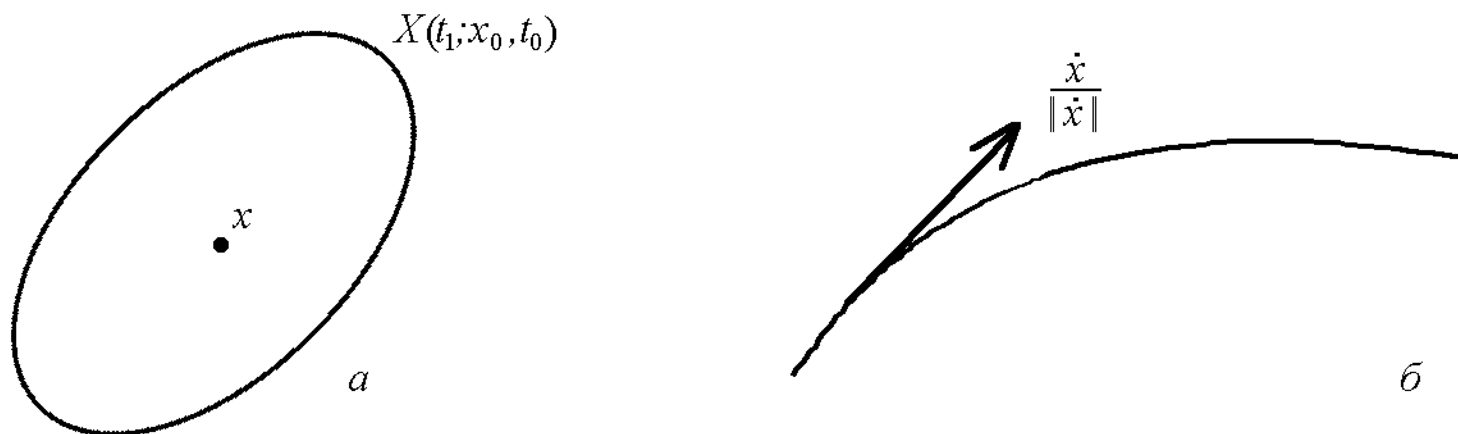


Рис. 11.1. Фронт волны возможностей (*а*) и движение по траекториям (*б*)

$\mathbf{x} \in \mathcal{R}_n$ в направлении $\mathbf{p} \in \mathcal{R}_n$, $||\mathbf{p}|| = 1$, может происходить с мгновенной скоростью $v(\mathbf{x}, \mathbf{p})$, возможность которой задана распределением $\mu_v(z, \mathbf{x}, \mathbf{p})$; если $\mu_v(z, \mathbf{x}, \mathbf{p}) = \mu_0$, то возможность равенства $v(\mathbf{x}, \mathbf{p}) = z$ равна μ_0; в этих равенствах векторы $\mathbf{x}$ и $\mathbf{p}$ рассматриваются как параметры.

Рассмотрим сначала случай, когда для каждого $\mathbf{x} \in \mathcal{R}_n$ возможно изменение состояния системы в любом направлении $\mathbf{p} \in \mathcal{R}_n$, но величина мгновенной скорости изменения для любых $\mathbf{x}$ и $\mathbf{p}$ однозначно определена:

$$\mu_v(z, \mathbf{x}, \mathbf{p}) = \begin{cases} 1, & \text{если } z = v_0(\mathbf{x}, \mathbf{p}), \\ 0, & \text{если } z \neq v_0(\mathbf{x}, \mathbf{p}). \end{cases} \tag{11.40}$$

Обозначим $\mathbf{X}(t_1; \mathbf{x}_0, t_0) \subset \mathcal{R}_n$ множество тех состояний системы, для которых минимальное время $t_{\min}(\mathbf{x}_1, \mathbf{x}_0)$ движения от $\mathbf{x}_0$ до любой точки $\mathbf{x}_1 \in \mathbf{X}(t_1; \mathbf{x}_0, t_0)$ не превосходит $t_1 - t_0$. Граница множества $\mathbf{X}(t_1; \mathbf{x}_0, t_0)$ является "фронтом" распространяющейся волны возможности. Положение фронта зависит от поля мгновенных скоростей и при указанных условиях однозначно им определяется, рис. 11.1.

Если распределение возможностей величины мгновенной скорости $v(\mathbf{x}, \mathbf{p})$ нечетко, то связь между $v(\mathbf{x}, \mathbf{p})$ и $t_{\min}(\mathbf{x}_1, \mathbf{x}_0)$ индуцирует распределение возможностей для величины $t_{\min}(\mathbf{x}_1, \mathbf{x}_0)$ и для "фронта" возможностей $\mathbf{X}(t_1; \mathbf{x}_0, t_0)$.

Найдем $t_{\min}(\mathbf{x}_1, \mathbf{x}_0)$ для случая, когда распределение мгновенной скорости задано в виде (11.40). Обозначим $f_0(\mathbf{x}, \mathbf{p}) = 1/v_0(\mathbf{x}, \mathbf{p})$ — "обратную" мгновенную скорость изменения состояния $\mathbf{x}$ системы в направлении $\mathbf{p}$, и рассмотрим линию в $\mathcal{R}_n$, соединяющую точки $\mathbf{x}_0$ и $\mathbf{x}_1$, заданную в параметрическом виде: $\mathbf{x} = \mathbf{x}(s)$, $s \in [s_0, s_1]$, причем $\mathbf{x}(s_0) = \mathbf{x}_0$, $\mathbf{x}(s_1) = \mathbf{x}_1$. Будем считать, что $\mathbf{x}(\cdot)$ дважды дифференцируема на $[s_0, s_1]$, а производную по параметру будем обозначать $d\mathbf{x}/ds = \dot{\mathbf{x}}$. Время движения по элементу длины $dl = ||\dot{x}||ds$ равно $f_0(\mathbf{x}, \mathbf{p})||\dot{\mathbf{x}}||ds$, где $\mathbf{p} = \dot{\mathbf{x}}/||\dot{\mathbf{x}}||$ — единичный вектор в направлении касательной к линии $\mathbf{x}(\cdot)$ в точке $\mathbf{x} = \mathbf{x}(s)$. Обозначим $f_0\left(\mathbf{x}, \dot{\mathbf{x}}/||\dot{\mathbf{x}}||\right) =$

$= f(\mathbf{x}, \dot{\mathbf{x}})$, тогда время движения от $\mathbf{x}_0$ до $\mathbf{x}_1$ по траектории $\mathbf{x}(s)$ равно

$$\tau(\mathbf{x}_1, \mathbf{x}_0) = \int\limits_{s_0}^{s_1} f(\mathbf{x}, \dot{\mathbf{x}}) \|\dot{\mathbf{x}}\| ds, \tag{11.41}$$

и

$$t_{\min}(\mathbf{x}_1, \mathbf{x}_0) = \inf\left\{\tau(\mathbf{x}_1, \mathbf{x}_0)\right\}, \tag{11.42}$$

где точная нижняя грань берется по всем траекториям, соединяющим $\mathbf{x}_0$ и $\mathbf{x}_1$. Уравнение Эйлера для задачи (11.42) имеет вид

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{x}}}\right) - \frac{\partial L}{\partial \mathbf{x}} = 0,$$

с граничными условиями

$$\mathbf{x}(s_0) = \mathbf{x}_0, \quad \mathbf{x}(s_1) = \mathbf{x}_1.$$

Здесь $L(\mathbf{x}, \dot{\mathbf{x}}) = f(\mathbf{x}, \dot{\mathbf{x}})\|\dot{\mathbf{x}}\|$ — функция Лагранжа задачи (11.42).

Рассмотрим, например, случай, когда $f(\mathbf{x}, \dot{\mathbf{x}})$ не зависит от $\mathbf{x}$. Введем для вектора $\dot{\mathbf{x}}$ сферическую систему координат и запишем уравнение движения $\dot{x}_i = \rho\vartheta_i(\varphi_1, \ldots, \varphi_{n-1})$, $i = 1, \ldots, n$; здесь $\rho = \|\dot{\mathbf{x}}\|$, а $\varphi = (\varphi_1, \ldots, \varphi_{n-1}) \in \mathcal{R}_{n-1}$ — угловые координаты, определяющие положение точки на n-мерной единичной сфере. Тогда $f(\mathbf{x}, \dot{\mathbf{x}}) = \Theta(\varphi_1, \ldots, \varphi_{n-1})$. Считая, что $\Theta(\cdot, \ldots, \cdot)$ имеет положительно определенную вторую производную $\Theta''_{\varphi\varphi} \in (\mathcal{R}_{n-1} \to \mathcal{R}_{n-1})$, запишем уравнения Эйлера в виде

$$\frac{d}{ds}\Theta(\varphi_1, \ldots, \varphi_{n-1}) = 0, \tag{11.43}$$

$$\frac{d}{ds}\left(\rho\frac{\partial \Theta}{\partial \varphi_i}\right) = 0, \quad i = 1, \ldots, n-1. \tag{11.44}$$

Из (11.43) имеем $(\partial\Theta/\partial\varphi)\dot{\varphi} = 0$, т.е. $\dot{\varphi}(s) = \mathbf{e}(s)$, где $\mathbf{e}(s)$ — любой вектор из $\mathcal{R}_{n-1}$, ортогональный Θ'_{φ}. Из (11.44) получаем равенство $\Theta'_{\varphi}\dot{\rho} + \rho\Theta''_{\varphi\varphi}\mathbf{e} = 0$, умножив которое скалярно на $\mathbf{e}(s)$, убедимся, что $\Theta''_{\varphi\varphi}\mathbf{e} = 0$. Значит, решением (11.43), (11.44) является $\rho = \mathrm{const}$, $\varphi = \mathrm{const}$, что с учетом граничных условий для траектории дает экстремаль функционала в (11.42) в виде

$$\mathbf{x} = (\mathbf{x} - \mathbf{x}_0)\frac{s - s_0}{s_1 - s_0} + \mathbf{x}_0.$$

Итак, экстремалями функционала в (11.42) являются отрезки прямой линии, соединяющие точки $\mathbf{x}_0$ и $\mathbf{x}_1$, и время движения по экстремали равно

$$t_{\min}(\mathbf{x}_1, \mathbf{x}_0) = \|\mathbf{x}_1 - \mathbf{x}_0\| \, f\left(\frac{\mathbf{x}_1 - \mathbf{x}_0}{\|\mathbf{x}_1 - \mathbf{x}_0\|}\right);$$

скорость изменения состояния вдоль экстремали постоянна.

Поверхность в $\mathcal{R}_n$ вида $\mathbf{x}_0 + v(\mathbf{p})(t_1 - t_0)\mathbf{p}$, где вектор $\mathbf{p} \in \mathcal{R}_n$ пробегает единичную сферу, дает фронт возможных состояний к моменту времени t_1, если в момент времени t_0 система находилась в состоянии $\mathbf{x}_0$. Движение этого фронта с течением времени t_1 можно интерпретировать как распространение волны возможностей.

Пусть теперь при всех $\mathbf{x} \in \mathcal{R}_n$ мгновенная скорость изменения состояния $v(\mathbf{p})$ в направлении $\mathbf{p}$ есть нечеткая переменная с распределением $\mu_{v(\mathbf{p})}(z) = \varphi_v(z, \mathbf{p})$. Тогда $t_{\min}(\mathbf{x}_1, \mathbf{x}_0)$ есть нечеткая переменная с распределением

$$\mu_t(\tau, \mathbf{x}_1, \mathbf{x}_0) = \varphi_v\left(\frac{\|\mathbf{x}_1 - \mathbf{x}_0\|}{\tau}, \frac{\mathbf{x}_1 - \mathbf{x}_0}{\|\mathbf{x}_1 - \mathbf{x}_0\|}\right),$$

и окончательно, переходная возможность равна

$$\mu(\mathbf{x}_1, t_1 | \mathbf{x}_0, \vartheta_0) = \varphi_v\left(\frac{\|\mathbf{x}_1 - \mathbf{x}_0\|}{\tau}, \frac{\mathbf{x}_1 - \mathbf{x}_0}{\|\mathbf{x}_1 - \mathbf{x}_0\|}\right).$$

В общем же случае, когда мгновенная скорость $v(\mathbf{x}, \mathbf{p})$ изменения состояния системы является нечеткой величиной с распределением $\varphi_v(z, \mathbf{x}, \mathbf{p})$, определим функцию

$$v(\lambda, \mathbf{x}, \mathbf{p}) = \sup\{z | \varphi_v(z, \mathbf{x}, \mathbf{p}) \leq \lambda\},$$

равную мгновенной скорости изменения состояния системы в точке $\mathbf{x}$ в направлении $\mathbf{p}$, возможность которой не превосходит λ, и вычислим $t_{\min}(\lambda, \mathbf{x}_1, \mathbf{x}_0)$ — минимальное время перехода системы из состояния $\mathbf{x}_0$ в $\mathbf{x}_1$, вычисляемое как значение функционала (11.41) на своей экстремали при $f(\mathbf{x}, \mathbf{p}) = 1/v(\lambda, \mathbf{x}, \mathbf{p})$. Это — наименьшее время перехода, возможность которого не больше, чем λ. В этом случае

$$\mu(\mathbf{x}_1, t_1 | \mathbf{x}_0, \vartheta_0) = \sup\{\lambda | t(\lambda, \mathbf{x}_1, \mathbf{x}_0) \geq t_1 - t_0). \tag{11.45}$$

По построению ясно, что $t(\lambda, \mathbf{x}_1, \mathbf{x}_0)$ монотонно не возрастает по λ, поэтому неравенство в (11.45) можно заменить на равенство.

11.8.7. Нечеткие процессы и наблюдения. При детерминированном моделировании динамических систем результат решения эволюционных уравнений дает однозначный ответ о значении

координат системы в тот или иной момент времени. При стохастическом моделировании его результат также определяет реальную наблюдаемую величину — частоту появления некоторого состояния системы. Результат же нечеткого моделирования говорит лишь о том, какие состояния системы более возможны в заданный момент времени, а какие — менее. Этот подход характеризуется определенным произволом в выборе шкалы возможностей: конкретные значения возможностей того или иного события используются лишь для того, чтобы упорядочить события, причем этот порядок не изменится при строго монотонном преобразовании шкалы возможности. В частности, если возможность некоторого значения координат системы в момент времени t_0 равна единице, это вовсе не означает, что такая ситуация действительно произойдет в реальности.

Однако, наблюдение состояния нечеткой системы в некоторый момент времени приводит к "редукции возможности", аналогичной редукции волнового пакета в квантовой механике, и в определенном смысле к уточнению описания динамической системы.

Пусть $\xi \in \mathcal{R}_n$ — результат измерения координаты нечеткого процесса $\eta(t)$, проведенного в момент времени $t \in [0, T)$ по схеме

$$\xi = \eta(t) + \nu, \tag{11.46}$$

где $\eta(t) \in \mathcal{R}_n$, $\nu \in \mathcal{R}_n$ — нечеткая погрешность измерения с распределением возможностей $\mu_\nu(\cdot)$. Найдем, в чем выражается редукция возможности при нечетком наблюдении координат системы. Пусть распределение возможностей нечеткого элемента $\eta(t)$ при заданном t определяется функцией $\mu_{\eta(t)}(\cdot)$, и нечеткие элементы $\eta(t)$ и ν независимы. Тогда совместное распределение пары нечетких элементов (η, ν) задается функцией

$$\mu_{(\eta,\nu)}(x, y) = \min(\mu_{\eta(t)}(x), \mu_\nu(y)), \quad x, y \in \mathcal{R}_n,$$

а совместное распределение нечетких элементов $\eta(t)$ и ξ — функцией

$$\mu_{(\eta(t),\xi)}(x, z) = \min\{\mu_{\eta(t)}(x), \mu_\nu(z - x)\}, \quad x, z \in \mathcal{R}_n.$$

Маргинальное распределение нечеткого элемента ξ равно

$$\mu_\xi(z) = \sup_{x \in \mathcal{R}_n} \min\{\mu_{\eta(t)}(x), \mu_\nu(z - x)\},$$

а условное распределение $\eta(t)$ при наблюдении ξ определяется соотношением

$$\mu_{(\eta(t)|\xi)}(x|z) \begin{cases} = \mu_{(\eta(t),\xi)}(x, z), & \mu_\xi(z) > \mu_{(\eta(t),\xi)}(x, z), \\ \geq \mu_{(\eta(t),\xi)}(x, z), & \mu_\xi(z) = \mu_{(\eta(t),\xi)}(x, z), \end{cases} \tag{11.47}$$

Вариант условного распределения далее используется как начальное распределение возможностей координат системы для начального момента времени $t_0 = t$.

Заметим, что самое “четкое” утверждение о значении некоторой переменной состоит в следующем: “значение z переменной вполне возможно, а остальные — совершенно невозможны”. В терминах теории возможностей это означает, что распределение возможности для этой переменной равно нулю всюду, кроме точки z. Это соображение лежит в основе введенного ранее, в разделе , сравнения по “четкости” двух нечетких элементов: считается, что η_1 не менее четкий, чем η_2, если для распределений их возможностей выполнено неравенство $\mu_{\eta_1}(z) \leq \mu_{\eta_2(z)}$ для всех z. (На языке теории нечетких множеств такое равенство означает, что множество с характеристической функцией $\mu_{\eta_1}(z)$ является подмножеством множества $\mu_{\eta_2}(z)$). Сравнение по четкости для нечетких элементов сохраняется при монотонном преобразовании шкалы возможности, так же, как и свойство одного множества быть подмножеством другого.

В указанном смысле распределение возможностей (11.47) — не менее четкое, чем $\mu(x,t)$, так как $\mu_{\eta(t)|\xi}(x,t|y) \leq \mu(x,t)$. Величина

$$\sup_x \mu_{\eta(t)|\xi}(x,t|y)$$

характеризует согласие нечеткой модели динамической системы с результатом эксперимента, так как она равна возможности описания системы заданной моделью при условии, что в результате измерения координат системы в момент времени t получено значение y. В частности, если измерение (11.46) проведено безошибочно, т.е. ν в (11.46) “четко” и равно нулю, то в (11.47) следует положить

$$\mu_\nu(z) = \begin{cases} 0 & \text{если } z \neq 0 \\ 1 & \text{если } z = 0 \end{cases}, \quad z \in \mathcal{R}_n.$$

Предложенная модель описания эволюции динамической системы в условиях неопределенности, основанная на конструкциях теории возможности, следует той же схеме, по которой построена теория случайных процессов. Описание динамической системы получено в терминах возможности тех или иных ее состояний; согласие модели с реальностью допускает проверку на основании эксперимента, результат которого может быть как вполне определен (“четкий” случай), так и не вполне определен (нечеток). Описание инвариантно по отношению к преобразованиям шкалы значений возможности, сохраняющей порядок.

Такой подход в ситуациях, не описывающихся стохастическими моделями, позволяет указывать, какие состояния системы более возможны, а какие — менее.

11.8.8. Распространение возможностей. В предыдущих пунктах 11.8.6, 11.8.7 рассмотрен пример, когда система в каждый момент времени характеризуется вполне определенными значениями координат, но доступная исследователю информация об этих значениях представлена только в виде распределения возможностей $\mu(\mathbf{x}, t)$.

Иная ситуация, которая тоже приводит к распространению волн возможностей, возникает, например, при описании возможности возгорания в данной точке пространства при распространении пожара, возможности разрушения при распространении селей, наводнений и т.п., возможности заболевания при распространении эпидемий и т.п. Так же, как и в предыдущем случае, задается начальное распределение возможностей наступления данного события A (пожара, эпидемии, разрушения и т.п.) на множестве точек выделенной области X пространства $\mathcal{R}_n$ и мгновенная скорость его распространения. Теперь, однако, дополнительная информация о том, что в момент времени $t \in [0, T]$ в точке x событие A наступило, не приводит к редукции фронта распространения волны возможностей.

Пусть на области $X \subset \mathcal{R}_2$ задано начальное распределение $\mu(\mathbf{x}, 0)$ возможностей того, что в точке $\mathbf{x} \in X$ в момент $t = 0$ наступило событие A, и известно, что мгновенная скорость $v(\mathbf{x}, \mathbf{p})$ распространения события A является нечеткой величиной с распределением $\varphi_v(z, \mathbf{x}, \mathbf{p})$. Тогда распределение возможностей $\mu(\mathbf{x}, t)$ наступления события A в каждой точке области X в момент времени t вычисляется методами, описанными в разделе . Если дополнительно из независимых источников информации известно распределение $\mu_{add}(\mathbf{x}, \vartheta_0)$ возможностей наступления A в момент времени $t_0 \in [0, T]$, то учет этой информации дает в результате распределение

$$\mu_0(\mathbf{x}, t_0) = \max\{\mu(\mathbf{x}, \vartheta_0), \mu_{add}(\mathbf{x}, \vartheta_0)\}.$$

Это распределение считается начальным для моделирования дальнейшего распространения возможностей наступления события A в точках области X.

Список литературы

1. *Анищенко В.С.* Сложные колебания в простых системах. — М.: Наука, 1990.
2. *Арнольд В.И.* Дополнительные главы теории обыкновенных дифференциальных уравнений. — М.: Наука, 1978.
3. *Боголюбов Н.Н., Митропольский Ю.А.* Асимптотические методы в теории нелинейных колебаний. — М.: Наука, 1974.
4. *Васильева А.Б., Бутузов В.Ф.* Асимптотические разложения решений сингулярных уравнений. — М.: Наука, 1973.
5. *Гилмор Р.* Прикладная теория катастроф. — М.: Мир, 1984.
6. *Дымников В.П., Филатов А.Н.* Основы математической теории климата. — М.: ВИНИТИ, 1994.
7. *Заславский Г.М., Сагдеев Р.З.* Введение в нелинейную физику. — М.: Наука, 1988.
8. *Лоскутов А.Ю., Михайлов А.С.* Введение в синергетику. — М.: Наука, 1990.
9. *Климонтович Ю.Л.* Турбулентное движение и структура хаоса. — М.: Наука, 1990.
10. *Петровский И.Г.* Лекции по теории обыкновенных дифференциальных уравнений. — М.: Изд-во Моск. ун-та, 1984.
11. *Пригожин И., Стенгерс И.* Время, хаос, квант: к решению парадокса времени. — М.: Прогресс, 1994.
12. *Пытьев Ю.П.* Методы анализа и интерпретации эксперимента. — М.: Изд-во Моск. ун-та, 1990.
13. *Пытьев Ю.П.* Математические методы интерпретации эксперимента. — М.: Высшая школа, 1989.
14. *Пытьев Ю.П., Шишмарев И.А.* Курс теории вероятностей и математической статистики для физиков. — М.: Изд-во Моск. ун-та, 1983.

15. *Пытьев Ю.П.* Морфологический анализ изображений. // Докл. АН СССР. — 1983. — Т.26, № 5. — С.1061–1064.

16. *Пытьев Ю.П.* Возможность. Элементы теории и применения. — М.: Эдиториал УРСС, 2000.

17. *Пытьев Ю.П., Чуличков А.И.* ЭВМ анализирует форму изображения. — М.: Знание, 1988. (Сер. “Математика. Кибернетика”, № 5.)

18. *Романовский Ю.М., Степанова Н.В., Чернавский Д.С.* Математическая биофизика. — М.: Наука, 1984.

19. *Хакен Г.* Синергетика. Иерархии неустойчивостей в самоорганизующихся системах и устройствах. — М.: Мир, 1985

20. *Шустер П.* Динамический хаос. — М.: Мир, 1988.

Научное издание

ЧУЛИЧКОВ Алексей Иванович

МАТЕМАТИЧЕСКИЕ МОДЕЛИ НЕЛИНЕЙНОЙ ДИНАМИКИ

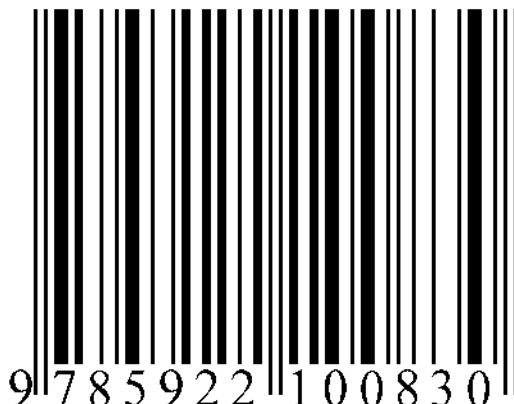
Редактор *Е.Ю. Ходан*
Оригинал-макет автора

ЛР № 071930 от 06.07.99. Подписано в печать 10.10.2000.
Формат 60×90/16. Бумага типографская. Печать офсетная.
Усл. печ. л. 18,5. Уч.-изд. л. 20,35. Тираж 400 экз. Заказ №

Издательская фирма "ФИЗМАТЛИТ"
МАИК "Наука/Интерпериодика"
117864 Москва, ул. Профсоюзная, 90

Отпечатано с готовых диапозитивов в Московской типографии № 6
Министерства Российской Федерации по делам печати,
телерадиовещания и средств массовых коммуникаций.
109088 Москва Ж-88, Южнопортовая ул., 24

ISBN 5-9221-0083-1



A.I. CHULICHKOV

MATHEMATICAL MODELS OF NONLINEAR DYNAMICS

PHYSICS AND MATHEMATICS PUBLISHERS

International Academic Publishing Company "Nauka"
Russian Academy of Science

Moscow, 2000, 296 pages

Abstract. In this book the known methods of mathematical modelling of dynamic systems are generalised and new methods are offered. The survey presentation of known methods of modelling from unified positions is directed on explaining the structure of a characteristic behaviour of complex non-linear systems, in particular, the phenomenon of self-organising. Entirely new approach to modelling of dynamic systems is based on the theory of capabilities and fuzzy mathematics and is oriented on the description of dynamics in conditions of uncertainty, in situations, when the stochastic modeling is ineffective. At fuzzy modelling it is possible to estimate only certain preferences in a system behaviour by setting the order on the state space indicative of what behaviour of a system is more possible, and what less.

Readership. Engineers and specialists in mathematical modeling, undergraduate and postgraduate students of the universities and technical schools.

Contents. Quality analysis of dynamic systems, mathematical models of chaotic dynamics, stochastic and fuzzy models.

Author. Chulichkov Alexey Ivanovich (b. 1954), D.Sc. (Phys., Math., 1993), professor of the Moscow State University named after M.V.Lomonosov, Physics dept. One of the leading specialist in Russia in mathematical modeling of physical systems and measuring processes; in mathematical aspects of the problems of the measuring analysis, planning and interpretation; in image processing.

Библиография

[1] Анищенко В.С. Сложные колебания в простых системах. – М.: Наука, 1990.

[2] Арнольд В.И. Дополнительные главы теории обыкновенных дифференциальных уравнений. - М.: Наука, 1978.

[3] Боголюбов Н.Н., Митропольский Ю.А. Асимптотические методы в теории нелинейных колебаний. - М.: Наука, 1974.

[4] Васильева А.Б., Бутузов В.Ф. Асимптотические разложения решений сингулярных уравнений. – М.: Наука, 1973.

[5] Гилмор Р. Прикладная теория катастроф. - М.: Мир, 1984.

[6] Дымников В.П., Филатов А.Н. Основы математической теории климата. - М.: ВИНИТИ, 1994.

[7] Заславский Г.М., Сагдеев Р.З. Введение в нелинейную физику. - М.: Наука, 1988.

[8] Лоскутов А.Ю., Михайлов А.С. Введение в синергетику. - М.: Наука, 1990.

[9] Климонтович Ю.Л. Турбулентное движение и структура хаоса. - М.: Наука, 1990.

[10] Петровский И.Г. Лекции по теории обыкновенных дифференциальных уравнений. – М.: Изд-во Моск. ун-та, 1984.

[11] Пригожин И., Стенгерс И. Время, хаос, квант: к решению парадокса времени. – М.: Прогресс, 1994.

[12] Пытьев Ю.П. Методы анализа и интерпретации эксперимента. – М.: Изд-во Моск. ун-та, 1990.

[13] Пытьев Ю.П. Математические методы интерпретации эксперимента. – М.: Высшая школа, 1989.

[14] Пытьев Ю.П., Шишмарев И.А. Курс теории вероятностей и математической статистики для физиков. – М.: Изд-во Моск. Ун-та, 1983.

[15] Пытьев Ю.П. Морфологический анализ изображений. — Докл. АН СССР, 1983. Т.26, N 5. С.1061-1064.

[16] Пытьев Ю.П. Основы теории возможностей. Методы оптимального оценивания и принятия решений. – Вестник Моск. ун-та. Сер.3 Физика. Астрономия. 1997. Т. 38, 1, 5, 6.

[17] Пытьев Ю.П., Чуличков А.И. ЭВМ анализирует форму изображения. -. Изд-во "Знание". Серия "Математика. Кибернетика". 1988г., N 5.

[18] Романовский Ю.М., Степанова Н.В., Чернавский Д.С. Математическая биофизика. - М.: Наука, 1984.

[19] Хакен Г. Синергетика. Иерархии неустойчивостей в самоорганизующихся системах и устройствах. - М.: Мир. 1985

[20] Шустер П. Динамический хаос. - М.: Мир, 1988.